DOCUMENT RESUME

ED 223 512	SO 014 382

AUTHOR	Colby, Anne; And Others
TITLE	A Longitudinal Study of Moral Judgment.
INSTITUTION	Harvard Univ., Cambridge, Mass. Graduate School of Education.
PUB DATE	83
NOTE	198p.; For a related document, see ED 172 946 and SO 014 384. Forthcoming in: Monograph of the Society of the Research in Child Development, 1983.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Chronological Age; *Developmental Stages; Educational Background; Intelligence Quotient; Longitudinal Studies; Males; *Moral Development; Moral Values; *Research Design; *Research Methodology; Socioeconomic Influences; *Test Reliability; *Test Validity; Value Judgment
IDENTIFIERS	Kohlberg (Lawrence)

ABSTRACT

A 20-year study to verify Lawrence Kohlberg's theory of moral development through a new research design, the Standard Issue Scoring System, is reported. Kohlberg theorizes that an individual progresses through several stages in attaining moral judgment. As children grow older, they are able to integrate diverse points of view on a moral conflict. Thus, each stage presupposes understanding gained at previous stages. The Standard Issue Scoring System for monitoring developmental stages replaces earlier evaluation techniques which have not proved reliable. A major portion of the document is devoted to a detailed description of the method, reliability, validity, and results of Standard Issue Scoring. The study found that the 50 males interviewed (6 times in 3- to 4-year intervals) proceeded through the stages in the hypothesized sequence. Moral judgment was found to be positively correlated with age, socioeconomic status, IQ, and education. The Standard Issue Scoring System was found to provide a valid measure. (KC)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

ED223512

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Lawrence
Kohlberg
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A Longitudinal Study of Moral Judgment

Anne Colby

Lawrence Kohlberg

John Gibbs

Marcus Lieberman

Monograph of the Society of the Research in Child Development, 1983, in press.

SO 014 382

Headings and Sub-Headings

- Introduction
  - Development of the Standard Issue Moral Judgment Interview and Scoring System
  - Moral Judgment Interview
  - The Explicated Criterion Judgment
  - Standard Issue Scoring Rules
  - Summary
- Method
  - Design
  - Subjects
  - Longitudinal Follow-up and Sample Attrition
  - Moral Judgment Interviews
  - Moral Judgment Scoring
- Reliability and Validity of Standard Issue Scoring
  - Reliability
    - Test-retest Reliability
    - Inter-rater Reliability
    - Alternate Form Reliability
  - Equivalence of Forms A, B, and C
  - Internal Consistency
  - Conclusions - Reliability of Standard Issue Scoring
  - Validity of the Standard Form
- Results
  - Invariant Stage Sequence
  - Internal Consistency
  - Factor Analyses

Headings and Sub-Headings (continued)

9

Results (continued)
    Relations of moral judgment stage to age
    Cohort Effect
    Stability of Individual Differences
    The Relationship of Moral Judgment Stage to Other Variables
        Socioeconomic status
        Sociometric status
        Intelligence
        Education level attained
Discussion
    Standard Scoring Reliability and Validity
    Theoretical Evolution
    Age Norms
    Individual Differences
    Social Role and Status Variables and Moral Development
    Formal Education and Moral Development
    IQ and Moral Development
    College Relativism and "Regression" in the Longitudinal Sample
    The Stage Model

A Longitudinal Study of Moral Judgment

Abstract

This paper presents the results of a twenty-year longitudinal study of moral judgment development. The study represents an attempt to document the basic assumptions of Kohlberg's cognitive-developmental theory of moral judgment. Subjects were 50 boys aged 10, 13, and 16 at time 1 and were approximately equally divided at each age by social class and sociometric status. Sociometric and socioeconomic groups were equalized for intelligence. The study included six testing times - the original interview and five follow-up interviews administered at three to four year intervals. At each testing time subjects were individually interviewed on their judgments about nine hypothetical moral dilemmas. Interviews were stage scored according to the Standard Issue Scoring Manual (in press) - Forms A, B, and C. All scoring was done blind by individual dilemma. Data are presented on test-retest, alternate form, and inter-rater reliability for Standard Issue Scoring. Validity of the instrument is discussed.

It was found that subjects proceeded through the developmental stages in the hypothesized sequence. No subject skipped a stage in the sequence and only 4% (6) of the adjacent testing times showed downward stage change. This percentage was less than downward change on test-retest data. Moral judgment interviews also showed a high degree of internal consistency in stage scores assigned with the great majority of the interviews receiving all their scores at two adjacent stages. Factor analyses by dilemma and moral issue showed a single general moral stage factor.

Moral judgment was found to be positively correlated with age, sociometric status, IQ, and education. Stage scores in childhood were significantly

correlated with adulthood scores.

The results of this study were interpreted as being consistent with a cognitive developmental stage model. Subjects seemed to use a coherent structural orientation in thinking about a variety of moral dilemmas. Their thinking developed in a regular sequence of stages, neither skipping a stage nor reverting to use of a prior stage. The Standard Issue Scoring System was found to be reliable and it was concluded that it provides a valid measure of Kohlberg's moral judgment stages.

CONTENTS

Abstract 1

I. Introduction 5

II. Method 19

III. Reliability and Validity of Standard Issue Scoring 25

IV. Results 35

V. Discussion 51

Appendix A: Moral Judgment Interview 80

Appendix B: Illustrations of Moral Norms and Elements 93

Appendix C: Illustration of the Explicated Criterion Judgment 95

Appendix D: Standard Issue Scoring Rules 99

References 104

Acknowledgements 109

Footnotes 110

Tables 112

Figures 170

Chapter I

INTRODUCTION

This monograph presents the results of a twenty year longitudinal study of moral development. The study represents an attempt to document the basic assumptions of Kohlberg's cognitive developmental account of moral judgment. According to this account, moral judgment is said to develop through a sequence of six stages. Kohlberg (1969; 1976) followed Piaget (1965) in defining stages according to the following criteria:

1. Stages imply distinct or qualitative differences in children's modes of thinking or of solving the same problem at different ages.

2. These different modes of thought form an invariant sequence, order, or succession in individual development.

3. Each of these different and sequential modes of thought forms a "structured whole". A given stage response ... represents an underlying thought organization which determines responses to tasks which are not manifestly similar.

4. Cognitive stages are hierarchical integrations. Stages form an order of increasingly differentiated and integrated structures to fulfill a common function. (Kohlberg, 1969, pp 352-353).

In line with this stage notion, the cognitive developmental approach to moral judgment focuses on the qualitative form of the child's moral reasoning and on developmental changes in that reasoning. Kohlberg has attempted to describe general organizational or structural features of moral judgment that can be shown to develop in a regular sequence of stages. The concept of structure implies that a consistent logic or form of reasoning can be abstracted from the content of an individual's responses to a variety of situations. It implies that moral development may be defined in terms of the qualitative reorganization of the individual's pattern of thought rather than the learning of new content. Each new reorganization integrates within a broader perspective the insights that were achieved at lower stages. The developing child becomes better able to understand and integrate diverse points of view on a moral conflict situation and to take more of the relevant situational factors into account. In this sense, each stage presupposes the understanding gained at previous stages. As a result, each stage provides a more adequate way of making and justifying moral judgments. The order in which the stages develop is said to be the same in each individual not because the stages are innate, but because of the underlying logic of the sequence. (See Table 1 for a summary of the stages).

Insert Table 1 here

Kohlberg has hypothesized that the developmental levels that he has described constitute stages in a strict Piagetian sense. To test this hypothesis, longitudinal data are required following subjects over a relatively long span of time. First, the stage concept implies that under normal environmental conditions moral change will always be upward in direction. Second, it implies that there will be no stage skipping. The individual must pass through each stage in order to reach the next stage in the sequence. Third, the stage concept implies that individual's thinking will be at a single dominant stage across varying content, though usage of the stage adjacent to the dominant stage may also be expected to be present. Previous research has supported the general notion of an age developmental order of qualitative responses and a hierarchy of preference and comprehension in these levels. This has been shown by cross-sectional age studies in a variety of cultures (see Edwards, 1981). It has been shown by a number of training studies which support the assumption that change after exposure to moral judgments at other stages is always to the next stage up. For example, Blatt and Kohlberg (1975), Colby, Fenton et al, (1977), and Lockwood (1977) found that exposure to group moral discussions in which a range of stages was presented in general led to movement to the next stage up. Rest (1973) found that there was a Gutman scale hierarchy in comprehension of the stages, that is, that individuals comprehended all stages lower than their own dominant stage, they comprehended the next stage up if they exhibited on the pretest some (15%) usage of that stage, but they did not comprehend stages more than one above their own dominant stage.

In addition to data supporting the general idea of a developmental hierarchy in Kohlbergs' levels, a large number of studies reviewed by Blasi (1980) have generally found significant associations between moral judgment level and moral conduct. Reviews of the extant published literature on research in the development of moral judgment based on the Kohlberg levels generally support the cognitive-developmental assumptions as to the antecedents and correlates of moral judgment development (cf Rest, in press).

The data we have just cited as supporting the developmental level hypothesis have usually been data comparing group means. The results of these studies have not directly supported the strong stage claim, as critics like Kurtines and Grief (1974) have pointed out. In part, the ambiguity in some of these findings has been due to the limited reliability of Kohlberg's 1958 method of interviewing and scoring moral stage. The longitudinal data published by

Holstein (1976), Kuhn (1976), and White et al (1978) have shown some anomalies in stage sequence. It has not been clear whether these anomalies represent a failure of fit of the strict stage model to moral judgment development or whether they have represented confusions in the conceptual definitions of the stages or problems in the reliability or validity of the measure. In an early report of his own longitudinal data over ten years, Kohlberg (Kohlberg and Kramer, 1969) while reporting some measurement problems leading to anomalies, stressed a genuine failure of the stage sequence hypothesis in the college years leading to so-called "sophomore retrogression" in development. In 1973, Kohlberg reinterpreted the anomalies as resulting not from retrogression but from incorrect conceptualization of stages as they appear in development after high school. He reported some clinical analyses of cases suggesting that college retrogressors were exhibiting a stage "4½" (having moved out of conventional morality but not yet into principled moral judgment) and proposed revised conceptual definitions of the fourth and fifth stages. In fact, such revision accounted for only a few of the anomalies reported by Kohlberg and Kramer. This implied that there were general problems in the reliability and validity of the stage criteria and scoring method.

The present monograph attempts to address the validity of the stage model as applied to longitudinal data and the associated problems of stage definition and measurement which this task has required.

It reports a reanalysis of 1956-1968 longitudinal data along with analysis of the subsequent data collected from the same subjects from 1968 through 1976. The current analysis involved the application of a new scoring method (Standard

Issue Scoring) based on a substantially revised account of the stages. One particularly noteworthy change in moral judgment stage scoring is the omission of Stage 6 from the current manual. Stage 6 was omitted partly because none of the interviews in the longitudinal sample seemed intuitively to be Stage 6, partly because the standard dilemmas are not ideal for differentiating between Stages 5 and 6. The question of whether Stage 6 should be included as a natural psychological stage in the moral development sequence will remain unresolved until research (using more appropriate moral dilemmas and interviewing techniques) is conducted with a special sample of people likely to have developed beyond Stage 5.

Because the scoring method is critical to the validity of our study, we will include a fairly extensive description of Standard Issue Scoring before proceeding to the Methods chapter.

Development of the Standard Issue Moral Judgment Interview and Scoring System

In the early 1970's Kohlberg and his colleagues began to redefine the central features of the moral judgment stages and to construct a more adequate assessment instrument. Seven of Kohlberg's longitudinal cases were used as the data base for this endeavor. It was expected that this process would yield a more precise and accurate picture of moral judgment development. For example, it was expected that when Kohlberg's remaining longitudinal cases were analyzed using the revised stage criteria, the data would more closely fit the core assumptions of cognitive-developmental theory than they did in Kramer's original analysis. Since the longitudinal interviews used to validate the reformulations were kept entirely separate from those used to generate them, this revision process avoids circularity by allowing a test of prediction to data other than those used to generate the scoring system.[2]

The early (1958) scoring systems, Sentence Rating and Global Story Rating, were based on what was essentially a content analysis. That is, both systems focused on what concerns a subject brought to bear in resolving a dilemma (for example, a concern for love between a husband and wife, the importance of obeying the law, a fear of punishment) and treated those concerns as indicators of developmental stage. These scoring systems yielded enough sequence anomalies in the 1968 analysis of the longitudinal data to warrant substantial revision of the stage definitions. This revision process resulted in a clearer differentiation of moral judgment structure from content. That is, more formal or abstract features of moral judgment were identified and formed the core of the new stage conceptions.

The basic developmental concept underlying the revised stage sequence is level of sociomoral perspective, the characteristic point of view from which the individual formulates moral judgments. In discussing level of sociomoral perspective, let us begin by saying that we believe the perspective-taking underlying the moral stages to be intrinsically moral in nature rather than a logical or social-cognitive structure applied to the moral domain. In this interpretation we follow Turiel (1979) and Damon (in press) in their contention that there are many types of perspective-taking, each of which develops separately as a result of experience in a particular domain. In this view, spatial, social, and moral perspective-taking are fundamentally different processes rather than applications of a single general structure to different content areas. That is, the form of spatial perspective-taking is intrinsically spatial, not moral or social, and the form of moral perspective-taking is intrinsically moral, not spatial or moral.

According to this account, the most basic and general defining feature of moral judgment development is development through the levels of sociomoral perspective. These levels, briefly described in Table 1, provide a general organization of moral judgment and serve to inform and unite other more specific moral concepts such as the nature of the morally right or good, the nature of moral reciprocity, of moral rules, rights, obligation or duty, of fairness, welfare consequences, and moral values such as obediance to authority, preservation of human life, and maintenance of contracts and affectional relations. Within each of these specific moral concepts, the form of developmental change is to some extent specific to the nature of the particular concept in question. However, the general moral perspective can be seen to underlie its more specific manifestations.

The change in stage conceptions meant that the moral concerns which had been considered indicative of stage in the early systems became content in relation to the newly identified structures. This reformulation of core structure implied a need to redefine the unit of analysis in scoring such that the concerns that were confused with stage in the 1958 systems became the units of content for stage analysis in the 1971 Structural Issue Scoring System. In the latter system, stage was assigned to material within each content unit on the basis of level of perspective.

Structural Issue Scoring proved to be a substantial advance over earlier systems. The new stage definitions focusing on level of sociomoral perspective not only yielded more orderly data, they also made possible a more convincing rationale for the internal logic of the moral judgment stage sequence.

There were a number of problems with this approach, however. The need to determine level of perspective in stage scoring an interview implied that the scoring unit must be large and that scoring criteria must be very general and abstract. This meant that scoring decisions were subjective and often unreliable. Moreover, the reliance of this scoring system on very general features of stage structure meant that findings of invariant longitudinal sequence and stage consistency across issues could be attributed to consistency in or a universal sequence of the general features rather than providing evidence for consistency or sequentiality of the detailed conceptual differentations included in more specific moral stage definitions.

The aim of Standard Issue Scoring was to overcome these limitations of Structural Issue Scoring--to achieve greater objectivity and reliability in scoring by specifying clear and concrete stage criteria and to define the developmental sequences of the specific "moral concepts" within each stage as well as the sequence of the global or general stage structures. The redefinition of the scoring unit was the key to achieving these aims. An understanding of the general framework of the Standard Issue Scoring System and of the procedures used to construct it is essential to an explanation of its validity. Let us turn, therefore, to a description of Standard Issue Scoring.

The procedure for construction of the Standard Issue Scoring System was designed to avoid the problem of circularity (theoretical verification as a self-fulfilling prophecy). With this in mind, seven cases[3] were selected from Kohlberg's longitudinal sample. These "construction cases," chosen at random from among those tested at all six time points, were assigned global stage scores based on intensive discussion and analysis using concepts from

Structural Issue Scoring. The responses to each dilemma were then classified into clearly defined scoring units, or "interview judgments." Each of these interview judgments formed the basis for a "criterion judgment" to be entered in the scoring manual. The stage score of each criterion judgement was assigned on the basis of the global score of the interview from which it was derived as well as a conceptual analysis of the idea it embodied. The criterion judgments generated by these seven construction cases were later used to score the remaining interviews in the longitudinal study through a process of matching interview material to criterion judgments in the manual. Those cases not in the construction sample comprised a blind sample which was not used at all until the scoring manual had been completed.

Moral Judgment Interview

Three forms of moral judgment interview were constructed. Each form consists of three hypothetical moral dilemmas, and each dilemma is followed by nine to twelve standardized probe questions designed to elicit justifications, elaborations, and clarifications of the subject's moral judgments. For each dilemma these questions focus on the two moral issues that were chosen to represent the central value conflict in that dilemma. For example, the familiar "Heinz Dilemma" (Should Heinz steal a drug to save his dying wife if the only druggist able to provide the drug insists on a high price that Heinz cannot affort to pay?) is represented in Standard Scoring as a conflict between the value of preserving life and the value of upholding the law. Life and law are the two standard issues for this dilemma and the probing questions are designed to elicit information on the subjects' conceptions of these two issues.

Of course, the dilemma can also be seen to involve other value conflicts, e.g. between love of the husband for his wife (affiliation) and the druggist's property rights (property). In order to standardize the set of issues scored for each interview, the central issues for each dilemma were predefined. This preidentification of standard issues not only allows the sampling of moral judgments about the same six issues (per interview) for each subject, it also allowed the construction of three parallel forms of the moral judgment interview. Thus, the first dilemma in interview forms A and B focuses on the same two issues, Life and Law. The second dilemmas of the two forms concern the conflict between morality/conscience (whether to be lenient toward someone who has broken the law out of conscience) and punishment (whether to punish someone who has broken the law). The third dilemmas involve a conflict between authority (e.g. obeying one's parent) and contract (abiding by or holding someone to an agreement) Form C involves the same six issues but in somewhat different pairs than in Forms A and B. (See Appendix A for interview forms.)

The first step in Standard Issue Scoring involves the classification of the subject's responses to a dilemma into the two standard issue categories. This is a fairly simple procedure. In the Heinz dilemma, for example, all responses arguing for stealing the drug are classified as upholding the life issue; all those arguing against stealing the drug are classified as upholding the law issue. In the follow-up dilemma (Should Heinz be punished if he does steal the drug?) all responses arguing for leniency are classified as morality/conscience; all those arguing for punishment are classified as punishment.

The first step in scoring responses to a dilemma, then, is to separate material into the two issue categories. Since the issue units are large and often contain a great deal of material, Standard Issue Scoring involves two further subdivisions (by norm and by element, see Table 2) before stage

Insert Table 2

scoring begins. This results in a fairly small unit of analysis. The Standard Scoring unit, the criterion judgment, is defined by the intersection of Dilemma x Issue x Norm x Element. The additional procedural complications of subdivision by norm and element were found to be necessary in order to define a unit that was narrow enough to be homogeneous, to capture what seems to be a single, discrete moral concept or idea, yet broad enough to represent the idea's full conceptual or structural significance for the subject. That is, the system provides a way for the scorer to categorize interview material in a nonarbitrary way into manageable, conceptually coherent units (interview judgments) which can then be stage scored by matching them to very specific and concrete criteria in the scoring manual (Criterion Judgments). (See Appendix B for illustration of norms and elements).

This system of content classification provides for a meaningful definition of the scoring unit which, in addition to being essential to those preparing the scoring manual, is also necessary for each scorer in the process of analyzing an interview. In effect, material in the interview transcript is classified according to three types of content category before it is classified by stage or structure.

In addition to resolving the unit problem, this approach is useful in preventing some of the content-structure confusions which have been problems for earlier moral judgment scoring systems. For example, Stage 3 reasoning often focuses on love or the affiliative relationship as a reason for Heinz to steal the drug in Dilemma III. That is, the content of affiliation is likely to occur in the context of a Stage 3 (interpersonal concordance) structure. Earlier scoring systems, which failed to clearly differentiate content and structure, tended to mis-score as Stage 3 reasoning that was in fact structurally more advanced but that focused on affiliative content. (For example, "Heinz should steal the drug for his wife because of his deep commitment to her and to the marriage and the responsibility that results from that commitment".) By first categorizing according to content and then addressing the questions of structure or stage, Standard Scoring procedures involve explicit differentiation of form and content and, in effect, remind the rater that identification of a particular content has not answered the stage scoring question.

The Explicated Criterion Judgment

A structural or theoretical explication has been provided for each criterion judgment in the manual. Each explication includes a statement of the underlying stage structure reflected in the criterion judgment (Stage Structure), detailed criteria for defining a match to the criterion judgment (Critical Indicators), explanations of distinctions among criterion judgments a scorer is likely to confuse, and several examples of interview material which can be considered to match the criterion judgment. (See Appendix C for illustration of explicated criterion judgment).

This criterion judgment format proved an important advance over earlier scoring systems. The subjectivity of scoring decisions is minimized by the concrete and explicit specification of exactly what constitutes a match between interview material and a criterion judgment in the scoring manual.

Standard Issue Scoring Rules

As we have said, Standard Issue Scoring involves first classifying the responses to each dilemma into two broad categories - the standard moral issues for that dilemma. Within each issue a stage score is entered for each match between a criterion judgment in the manual and a moral judgment in the interview. Usually, somewhere between one and five such matches are assigned for the issue. Although in practice these matches tend to cluster at a single stage or at two adjacent stages, there is no restriction in the scoring rules that requires such consistency. The rules allow scores to be assigned at all five stages if matches are found at those stages.

In calculating an overall score for a three-dilemma interview form, one assigns a summary score for each of the six issues. As many as three stages may be represented in the issue summary score. The six issue scores are then combined to yield a global interview score and a continuous "moral maturity" score (weighted average) for the interview. (See Appendix D for further discussion of scoring rules.)

Summary

In summary, Kohlberg has postulated six stages in the development of moral judgment and has described these stages as characterized by holistic internal organization or internal consistency of structure across differing content and by the invariant sequence through which individuals proceed through the stages. In Kohlberg's early work these stages were assessed using an interview based on hypothetical moral dilemmas and two systems for scoring responses to those interviews: Sentence Rating and Global Story Rating. Some twelve years later, a new scoring system, Structural Issue Scoring, was introduced. From a psychometric point of view, there were serious problems with these scoring systems. The early systems lacked validity as measures of structural-developmental stages since they involved what was essentially a content analysis of interview responses. The later system achieved increased validity when used by very experienced raters but was subjective, difficult to use, and unreliable, particularly when used by those without extensive training and study with Kohlberg. Data will be presented here to show that the current system, Standard Issue Scoring, provides for reliable and valid assesment of moral judgment stage. This was achieved through redefining the unit of analysis and the relation between moral judgment content and structure and through specifying more precisely the process of inference from interview material to stage scores. We will also argue that this improved method of analysis when applied to Kohlberg's longitudinal data provides a more adequate test of the stage hypothesis in moral judgment development than has been possible to date.

Chapter II

METHOD

Design

The design of the original cross-sectional study which was later followed up longitudinally was determined by a number of theoretical concerns. Three variables were included in the design: age, socioeconomic status, and sociometric status. Socioeconomic status was expected to be positively associated with moral judgment development in part because it was assumed to be an indicator of sense of participation in the society as a whole. Kohlberg, drawing on G. H. Mead (1934) considered this sense of participation to be an important determinant of moral development. Sociometric status was intended to be an indicator of peer group participation which Piaget (1965) argued was crucial to moral development. The age variable was intended to establish the age developmental characteristics of types of response to the moral dilemmas. These originally cross-sectional subjects, stratified by three levels of age and two levels of social class and sociometric status, were followed longitudinally at regular three to four year intervals for twenty years.

At each testing time subjects were interviewed on the nine hypothetical moral dilemmas making up the three forms of Kohlberg's moral judgment interview: Forms A, B, and C. At some testings subjects also responded to additional instruments not reported in this monograph including interviews concerning attitudes toward social and occupational roles and attitudes towards sex, the Thematic Apperception Test, Loevinger's Sentence Completion Test, Piagetian cognitive measures, and Selman's role-taking interview.[4]

Subjects

The basic sample consisted originally of eighty-four boys[5] filling the following 2x2x3 factorial design.

Insert Table 3

The population had the following characteristics:

a) Age - Subjects were 10, 13, and 16 at Time 1.

b) Class - To facilitate filling the design, two suburban Chicago school systems were selected, one predominantly upper-middle calss, the other predominantly lower-middle and working class. Fourth, seventh, and tenth grade classes formed the basis for selection. A dichotomous judgment of a boy's socioeconomic status was based on his parents' occupation and education, as reported in the school folder. In spite of efforts to obtain discrete groups, it was necessary to take children along a fairly broad continuum with a rather arbitrary, though conventional, dividing point. The fathers of boys in lower and lower-middle class group included unskilled, semi-skilled, and skilled laborers and white collar workers without a college education. The fathers of boys in the upper-middle class group included small businessmen, accountants, and salesmen with a college education, semi-professionals, executives, and professionals.

c) Sociometric status - When entering a given classroom the investigator described the procedures to be followed, including a "revealed differences" discussion among three boys. The boys were then asked to write the name of three other boys with whom they would like to have the discussion. The sociometric test was informally discussed with the teacher and compared with notes in the school folder before a final selection was made, in order to somewhat reduce determinants of school and athletic achievement and temporary fluctuations of popularity. The teachers were asked to comment on the boys' social connectedness, not on their moral characters or reputations.

As in the case of socioeconomic status, there were not enough subjects available to obtain only extreme groups, so that the dichotomy tends to divide a continuum. Boys who were never chosen or who were chosen once or twice but never by someone they had themselves chosen were designated sociometric isolates. Boys with at least two reciprocal choices or who were chosen at least three times, at least once reciprocally, were designated as integrates.

Intelligence - IQ scores were taken from school records at Time 1 and were based on various group tests routinely administered in the various schools (e.g. the Otis and the Thurstone P.M.A.). An attempt was made to equalize intelligence for the social class and sociometric groups. Those whose IQ's were above 120 or below 100 were excluded from the study. Although complete equalization was not achieved, the IQ differences were small and non-significant. Mean IQ for middle class boys was 109.7, for lower class 105.9. Mean IQ for integrates was 111.2, for isolates 104.4.

Religion and ethnicity - The ethnic and religious composition of our sample is presented in Table 4.

Insert Table 4

Longitudinal Follow-up and Sample Attrition

The study included six testing times - the original interview and five follow-up interviews. Because not every subject could be reached at each time point, the number of interviews per subject ranged from one to six. Only those subjects with at least two interviews were included in our current analysis. Fifty-three subjects met this condition. As shown in Table 5, the most typical number of interviews completed for a single subject was four. All but three subjects were interviewed at least three times. Age,

Insert Table 5

sociometric status, and socioeconomic status of origin of these 53 longitudinal subjects are presented in Table 3.

There was no attempt to locate for later longitudinal follow-up those boys who were unavailable for follow-up at Time 2. As shown in Table 3, the initial drop-out from the cross-sectional sample was not evenly divided by age and social class. Whereas 12 working class boys dropped out, only eight upper-middle class boys did so. Ten of the Time 1 16 year olds dropped out, while only three of the 10 year olds did so. The attrition of working class subjects was compensated for by adding to the sample in 1964 (Testing Time 3) five working class 19 year olds. Since IQ and sociometric status data are missing for these subjects (S's #91-96), they are not included in our analyses of social class, education, IQ and sociometric status in relation to moral judgment. However, since they were each interviewed at three or four time points, they were included in our analyses of longitudinal sequence and internal consistency of moral judgment. Thus the final longitudinal sample was skewed somewhat toward the younger cohort but was balanced in terms of socioeconomic status. The cohort bias should not affect the basic findings of the study, however, since orderliness of longitudinal sequence and internal consistency of responses to the interview were not related to age cohort.

In order to determine whether the subset of subjects who dropped out after Time 3 were higher or lower in their moral judgment than the sample as a whole, we compared Time 1 moral maturity scores of the whole sample, subdivided by age group, with Time 1 moral maturity scores of the drop-outs, again subdivided by age group. As shown in Table 6, the drop-out means were almost identical to the total sample means, and there was no tendency for subjects who dropped out earlier in the study to show lower scores at Time 1 than those who dropped out later. Thus, mean increase in moral maturity scores over time cannot be attributed to lower stage subjects dropping out of the longitudinal sample.

Insert Table 6

Moral Judgment Interviews

Subjects were interviewed first in 1955-1956 and at 3 to 4 year intervals thereafter. The last set of interviews (including 35 subjects) was completed in 1976-1977. Although the probe questions differed slightly from one testing time to another, the same nine dilemmas were used each time. These dilemmas were later used in constructing the Standard Issue Interviews and represent in effect the use of all three Standard Forms (A, B, and C) at each testing time for each subject. All interviews were conducted individually and were tape recorded and transcribed.

Moral Judgment Scoring

Interviews were scored according to the Standard Issue Scoring Manual (1983) Forms, A, B, and C. (See Appendix D for a discussion of scoring rules). Because the number of interviews to be scored was very large, a different rater scored each of the three forms A, B, and C. All three raters were highly experienced and reliability among them fell within the limits discussed in the reliability section of this monograph. None of the interviewers participated in coding the interviews. Scoring of all interviews was done blind. That is, raters coded the responses to each dilemma at each timepoint without knowledge of the subject's age or identity or of responses to (or scores on) other dilemmas at the same testing time or responses to (or scores on) any of the dilemmas at other testing times.

Chapter III

RELIABILITY AND VALIDITY OF STANDARD ISSUE SCORING

Reliability

Reliability data of several types were compiled for the Standard Issue instrument. As the following results indicate, the instrument has proven to be highly reliable.

Test-retest Reliability. Test-retest moral judgment interviews were conducted with 43 subjects using Form A, 31 subjects using Form B, and 10 subjects using both A and B. No test-retest data have been collected for Form C as yet. The same form(s) were used at Times 1 and 2 with conditions of testing held constant and intervals between Time 1 and Time 2 ranging from 3 to 6 weeks. Subjects were chosen from among volunteers in several Boston area elementary and high schools, colleges, and graduate schools. The college and graduate school students were paid for their time. Subjects ranged in age from 8 to 28 years and were approximately half male and half female.

Interviews were scored blind by two raters using the Standard Issue Scoring Manual (in press).

Test-retest reliability figures are summarized in Table 7. As shown in that table, correlations between Time 1 and Time 2 for Forms A and B are both in the high 90's. Since the correlations could be very high without much absolute agreement between scores at Time 1 and Time 2, we have also presented % agreement figures. For almost all subjects, the scores on Times 1 and 2 were within 1/3 stage of each other (One step - from 1 to 1(2) or 1(2) to 2(1) or 2(1) to 2, and so on.) If we look at global scores based on a nine point scale - the five stages and the four transition points between stages, we find between 70 and 80% complete agreement. We also calculated percent agreement using a more differentiated system of global scores with 13 categories. This system includes two transition points between each stage, distinguishing, for example, between an interview that is primarily Stage 2 with some 3 versus an interview that is primarily Stage 3 with some 2. The agreement levels for this system were in the 60's for the separate forms, 70% for the two forms combined.

Insert Table 7

Overall, then, it appears that on two interviews conducted about a month apart, almost all subjects receive scores within one third stage of each other. About three quarters receive identical scores on the two interviews when a nine point scale is used, and between one half and two thirds receive identical scores with the most finely differentiated 13 point scale. When scores do change from Time 1 to Time 2, the change is as likely to be negative as positive so it cannot be attributed to practice effect. There were no age or sex differences in test-retest stability.

Inter-rater Reliability. Test-retest interviews described above were also used for assessing inter-rater reliability. Twenty Form A interviews were scored independently by five raters. Ten Form B interviews were scored by four raters. In addition, 20 Form C longitudinal interviews were scored independently by two raters.

Percent agreement figures for inter-rater reliability on Form A ranged from 88 to 100% for agreement with a third of a stage, from 75 to 88% for complete agreement based on the nine point scale and from 53 to 63% for complete agreement using the most finely differentiated 13 point scale. The correlation for raters 1 and 2 on the Form A test-retest interviews was .98. Interrater figures for Forms B and C are about the same as those for Form A. (See Table 7).

Raters who scored these reliability data varied in degree of experience using the manual. Of the five Form A raters, two were familiar with Structural Issue Scoring as well as Standard Issue Scoring and in fact were the authors of Standard Issue Forms A and B (Colby and Gibbs). The third rater was also highly experienced but was not a manual author. The other two raters were much less experienced, having learned to score within the prior 6 to 8 months using the manual. They consulted with more experienced raters as they learned to use the manual to score a separate set of practise cases, but this consultation was no more extensive than would be provided at the week-long training workshop held at Harvard each year. Inter-rater reliability figures between the two "new scorers" and between each of them and an experienced scorer were at least as high as reliability among the experienced scorers. The same was true for Form B inter-rater reliability which involved two experienced and two new scorers. Thus it seems warranted to conclude that the Standard Issue Manual can be reliably mastered by relatively inexperienced users.

Insert Table 8

This conclusion is further supported by inter-rater reliability achieved between an experienced rater (Colby) and a group of research assistants from another university who had participated in a 4 day seminar on Standard Issue Scoring. On 10 of the interviews from a study being conducted by Norma Haan at University of California, Berkeley, percent agreement between Colby and the Berkeley raters was comparable to reliability figures achieved among Harvard raters. All pairs of scores were with one third stage of each other; 83%

agreed perfectly when a nine point scale was used.

In general, then, the figures for inter-rater reliability look roughly comparable to the test-retest figures - almost all interviews were scored within one third stage of each other by any two raters, and on about one half to two thirds of the interviews the two raters assigned identical scores even when using the 13 point scale.

Alternate Form Reliability. Alternate form data between Forms A and B are based on those 10 test-retest subjects who received both Forms A and B and on the 233 longitudinal interviews that included both forms. A single rater independently scored both forms of the 20 test-retest sample interviews. Percent agreement between Forms A and B for this sample was comparable to test-retest and inter-rater reliability: 100% of the interviews were given scores within one third stage of each other for the two forms, 75% received identical scores for A and B using the nine category system, and 67% received identical scores for the two forms using the 13 category system. The correlation between moral maturity scores for Forms A and B in this sample was .95.

The level of agreement across forms for the longitudinal data is not as high (see Table 9). This is to be expected since Form A was scored by rater 1 and Form B by rater 2. That is, the reliability figures confound form and rater differences. The correlation between Forms A and B for the longitudinal sample is .84.

Alternate form reliability between Form C and Forms A and B also represent conservative reliability estimates since they too confound form and rater effects. The correlation is .82 for Forms A and C, .84 for Forms B and C. That is, Form C is as highly correlated with each of the other two forms as they are with each other. Percent agreement figures are also comparable across all two form comparisons.

Insert Table 9

Since essentially unrelated developmental variables may be correlated if a wide age range is included, we have also computed alternate form correlations with age partialled out. The partial correlations are .78 for Forms A and B, .76 for A and C, and .78 for Forms B and C. Correlations between alternate forms were also computed within each age group. As shown in Table 10, the correlations are low, .37-.56, (probably due to restricted range) before age 16, but at age 16 and above they are in the .60's and .70's. In view of the relatively small n's, the fairly restricted MMS range even at the later ages, and the fact that each form was scored by a different rater, these correlations among the alternate forms can be considered psychometrically adequate.

Insert Table 10

Equivalence of Forms A, B and C

It has been demonstrated that there are high correlations among the three forms using the continuous moral maturity scores and a high correspondence between the qualitative global stage scores. However, there are small absolute differences between the forms. These differences affect the results of intervention studies with a pre- and post-testing design where one form is used as the pretest and the other as a post-test.

Table 11 reports the stage usage and moral maturity scores on each of the three forms at various ages. It can be seen that each form gives slightly different estimates of moral maturity than the others, and has slightly different pulls for certain stages, e.g. for Stage 1 and for Stage 5.

Insert Table 11

Because of the slight difference in "pull" between the forms, it is appropriate to provide a conversion formula or scale score which would equalize the three forms. Such a conversion formula adds a constant to the score on each form and a weighting factor for each form based on its correlations with the scores for the forms combined. This formula allows estimation of an absolute change value, given the use of one form for pretest and another for post-test. Readers should note that even with the creation of these conversion formulas it is usually preferable to counterbalance forms on change studies, e.g. giving half the subjects Form A on pretest and Form B on post-test and the other half Form B on pretest and Form A on post-test.

Regression analyses were performed to create scale scores based on the overall moral maturity score and using the score on each form as a single predictor variable.

For the sample of 199 interviews with all three forms used, Table 12 shows the coefficients and the proportion of variance accounted for by each form in predicting the overall score.

Insert Table 12

Regression was also performed on each age group, but the number of subjects ranged only from 20 to 40 and the results were highly unstable.

Internal Consistency

Cronbach's alpha was computed for each of the three interview forms, using issue scores as items. The results were as follows: Form A, .92; Form B, .96; and Form C, .94. These figures indicate that the measure meets the psychometric criterion of internal consistency of a test. That is, they provide strong evidence that the moral judgment interivew is measuring a single construct.

Recall in this regard that each of the three dilemmas within each form was scored independently.

Conclusions - Reliability of Standard Issue Scoring

A review of the correlational reliability data for Standard Issue Scoring indicates that the instrument is well within the limits of acceptable reliability. A comparison with related mesaures may be helpful here. Loevinger and Wessler (1970) report inter-rater reliability correlations for their Sentence Completion Test of ego development in the mid .80's (using total protocol scores) as compared to our .98. Rest (1979) reports test-retest reliability of .68-.92 and internal consistency reliability of .77 and .79 for the Defining Issues Test of moral development. Recall that test-retest reliability for Standard Scoring ranged from .96 to .99 and that internal consistency ranged from .92 to .96.

The qualitative analyses emphasized in many studies of moral development require more than high correlational reliability, however. Percent of absolute agreement between global scores is also important. In this regard the Standard Issue System again compares favorably with Loevinger's Sentence Completion Test. Inter-rater agreement on total protocol Sentence Completion scores using a ten point scale is reported to range from 50 to 80% (median 61%). This is substantially lower than the 75-88% for the nine point Standard Scoring scale and is, in fact, lower than even the 53-78% agreement on our thirteen point scale. Loevinger and Wessler find most protocols to receive scores within a half stage step of one another (88-100%, median-94%) while we find most protocols to be scored within one-third stage of each other (88-100%). Rest does not report percent agreement figures since his instrument does not yield stage typology scores.

In addition to test-retest, inter-rater, and alternate form reliability, we have also calculated the standard error of measurement for the moral judgment measure. With the standard error of measurement defined as $\sigma_{meas} = \sigma\sqrt{1-\rho_{xx}}$ (Nunnally, 1978), we have entered the standard deviation of the total longitudinal sample (69.87) as σ and .95 as ρ_{xx}, the reliability estimate of the measure. The standard error of measurement of the instrument based on these figures is 15.62 "moral maturity points".

Validity of the Standard Form

As we have argued elsewhere (Colby, 1978; Kohlberg, 1980), the appropriate validity concept for a developmental measure such as the Standard Issue system is construct validity, not prediction to an external criterion. For a measure of moral judgment stage, the two most critical empirical criteria of construct validity correspond to the two most central theoretical assumptions of the stage construct. They are invariance of stage sequence and "structural wholeness" or internal consistency (generality of stage usage across moral issues or dilemmas). As noted in the following description of our longitudinal data, the results confirm both invariant sequence and internal consistency. Among other things, we interpret "construct validity" to mean the fit of the data obtained by means of the test to primary components of its theoretical definition. The primary theoretical definition of structural moral development is that of an organization passing through an invariant developmental sequence. In other words, positive results of the longitudinal analysis support not only the theoretical assumptions, but also the validity of the measure. Negative results, of course, could be due either to an incorrect theory, an invalid test, or both. Furthermore, validity and reliability of a test are closely related notions since both refer to the generalizability of performance on a test, or a set of test items, to performance in other situations including the performance on other forms of the test or at other times of testing. In the case of structural stage, construct validity demands high generalizability or test-retest and alternate-form reliability. If a stage is a structural whole, the individual should be consistent over various stimuli and occasions of testing. Our reliability data fit this demand rather well.

In regard to the instrument's validity, one might also wonder whether Standard Issue Scores on responses to hypothetical dilemmas predict to moral judgment in real life. One should not expect, however, to find an exact correspondence between developmental stage of hypothetical and real moral judgments. As Damon (1977) has shown, the relation between moral judgments made by children in real and hypothetical versions of the same distributive justice situation is mediated by a number of factors including the individual's developmental level on the hypothetical judgments. Moreover, the most salient features of real-life dilemmas are not always unambiguously moral. Many of the issues dealt with in attempts to resolve such dilemmas may be practical or factual rather than moral. To the extent that such situations do elicit moral reasoning, the scores on individual's responses to hypothetical dilemmas should be predictive of, although not identical to, the scores on the real life moral judgments. Of course, evaulation of this aspect of construct validity is limited by the absence of techniques for scoring real-life moral judgments.

One study which has reported data on the relation between Standard Issue Scores and scores on real-life situations is Gilligan and Belenky's (1980) longitudinal study of women in the process of deciding whether or not to have an abortion. Gilligan and Belenky (1980) report rank order correlations on the relation between scores on the hypothetical abortion dilemma and the Standard dilemmas at Time 1 and Time 2 as .83 and .92, respectively. They present percent agreement figures but not correlations between scores on the real and hypothetical abortion dilemmas: 59% were scored at the same stage, all but one of the remainder were scored within one half stage of one another. Gilligan and Belenky do not, however, interpret the discrepancies between actual and hypothetical abortion dilemma scores as measurement error. Rather, they argue that the discrepancies are important psychological phenomena which have important implications for the future development and mental health of these women facing unwanted pregnancies. Like Selman and Jacquette (1978), they find that when individuals are unable to bring to bear their cognitive capacities (as measured in the hypothetical interview) in thinking about the real life situation, they are more likely to exhibit emotional disturbance both at the time of the initial interview and in longitudinal follow-up one year later. Gilligan and Belenky also report that changes in moral judgment scores on the Standard dilemmas in the year following the abortion were strong predictors of psychological adjustment at the time of the follow-up interview.

Chapter IV

RESULTS

Invariant Stage Sequence

According to our theoretical assumptions, the developmental sequence should be identical in every individual studied. Except under extreme circumstances, there should be no deviations from perfect sequentiality. That is, there should be no stage skipping and no downward stage movement. Let us first consider downward stage movement. Theoretically, our data should never yield a lower score at time n+1 than at time n. Since we have not constructed an error-free measure, however, some discrepant scores must be expected to result from measurement error. A reasonable estimate of the number of such deviations attributable to measurement error can be derived from analysis of test-retest reliability data. That is, we can assume that virtually none of our subjects will have changed in stage of moral judgment within the short test-retest interval. Any differences between scores at Times 1 and 2 can be attributed to measurement error. Therefore, the analysis of longitudinal sequence involved a comparison of the frequency of sequence reversals (downward stage movement) in the longitudinal data with the frequency of negative Time 1 - Time 2 changes in our test-retest data. (Of course, this comparison depends upon a psychometrically adequate level of test-retest reliability.) If sequence reversals exceed test-retest instability, we cannot consider our data to support the invariant sequence assumption. Table 13 presents global stage scores for the three interview forms A and B and C separately and combined. The sequence reversals are noted with an X. In Form A these sequence reversals (downward changes) occurred in only 7% of the adjacent time points using even our most differentiated 13 point scale of global interview scores. The reversals were 6% in Form B, 6% in Form C, and 5% in Forms A, B, and C combined. A comparison with downward stage change in test-retest data for Forms A and B is presented in Table 14. Since in every case the test-retest reversals are well over twice as great as the longitudinal reversals, it seems reasonable to attribute the violations of longitudinal sequence to measurement error.

Insert Tables 13 and 14

A second type of sequence analysis involves comparing the proportion of positive to negative stage change rather than focusing on negative change alone. Such a comparison is important because the invariant sequence assumption implies both that individuals will move upward through the developmental sequence and that they will not move downward. Data showing infrequent downward movement would be interpreted very differently if upward movement was equally infrequent. Although it is somewhat difficult to know what would constitute an adequate preponderance of upward stage change, Rest (1979) reports a ratio of 9.4 to 1 for positive to negative change on longitudinal data from the DIT. If we define change as one step of movement along our nine point scale, the ratio of positive to negative change in our study is 14.75 to 1.

In addition to predicting that at no time will a subject move downward in the developmental sequence, cognitive-developmental theory holds that each stage is a prerequisite for those that follow it. That is, the concept of invariant stage sequence implies that no stage will be omitted as development proceeds. Since within a four year interval a subject might enter and leave a stage, we could not necessarily expect our data to corroborate this aspect of sequentiality. Fortunately, however, the interval seems to have been short enough in relation to our subjects' rate of development to capture each stage in the sequence for each subject. In fact, Table 13 shows that in no case on any form did a subject reach a stage in the sequence without having gone through each preceding stage. For the most part, changes across the four year intervals were less than a full stage. In Forms A and B only 3% changed more than a full stage in four years. In Form C only 2% changed more than a full stage in four years. Scores combined across all three forms show only half a percent changing more than a full stage.

Internal Consistency

According to our theoretical assumptions, the logic of each stage forms a "structured whole". In line with this assumption, one would expect to find a high degree of internal consistency in stage scores assigned, at least within those units that are conceptually and psychologically coherent. The data support this assumption as clearly as they do that of invariant sequence.

One indication of degree of internal consistency in moral judgment is provided by distributions for each subject of proportion of reasoning scored at each of the five stages. Our analysis of these distributions showed that most interviews received all of their scores at either a single stage or at two adjacent stages. The mean percentage of reasoning at the individual's modal stage was 68% for Form A, 72% for Form B, 69% for Form C, and 67% for Forms A, B, and C combined. The mean percentage of reasoning at the subject's two most frequently used stages (always adjacent) was 98% for Form A, 97% for Form B, 99% for Form C, and 99% for Forms A, B, and C combined. (Remember that there are three dilemmas per form and each dilemma was scored without knowledge of responses to the other dilemmas, so these figures cannot be an artifact of scorer bias). The high correlations among alternate forms and high Cronbach's alpha figures reported in the section on reliability provide further support for the consistency of subjects' stage of reasoning across differing content.

Some interviews, however, received scores at three stages. In order to determine the number of interviews that did exhibit reasoning at three stages, it was necessary to establish a boundary below which the entry would be treated as error and above which it would be treated as real. For example, if an interview received 80% of its scores at Stage 1, 19% at Stage 2, and 1% at Stage 3, we would probably want to treat the 1% at Stage 3 as error and consider the interview to exhibit only two stages of moral judgment. We established the error boundary at 10% with entries of 10% and below treated as error, entries above 10% treated as real. Our choice of the 10% figure can be justified as follows. The 10% boundary derives from the relation between the number of criterion judgment matches assigned to an interview at a stage and the percentage of reasoning at that stage in the interview's distribution of stage scores. Through a comparison of distribution of stage use for each interview with a record of each individual CJ match score assigned, we determined that using a cut-off point of 10% would assure that in every case where the interview received one full criterion judgment match at a stage, the profile entry for that stage would be treated as real. That is, all profile entries of 10% and below reflected less than one criterion judgment match across the entire interview. For example, it might reflect the assignment of a transitional CJ score at that stage and the adjacent stage as in the case where a predominantly Stage 3/4 interview receives a single 2/3 score. Proportions of reasoning greater than 10% often reflected less than a full CJ on the score sheet, but proportions less than 10% never reflected a full CJ or more. Therefore our error boundary is conservative. Using the assessment of one full CJ as an indication of stage presence would have allowed us to treat many more entries as error. That is, our 10% boundary errs in the direction of treating as real some entries that are in fact error rather than in treating as error entries that are real. Even so, we find that only 9% of our longitudinal interviews show a third stage of reasoning greater than 10%. Using a less conservative cut-off, we find that only 1% of our interviews show a third stage of reasoning greater than 20%.

It is safe to say, then, that very few interviews show any usage of three stages of moral judgment. Where such three-stage spread does occur, it is always a matter of three adjacent stages. It is unclear, however, whether to interpret the instances where a third stage is used as cases of measurement error not caught by our conservative error boundary or as real but unusual cases of stage heterogeneity in the subject's thinking. In at least two cases, to be discussed in more detail later, we interpret the lack of consistency as a real phenomenon. These two cases, Subject 2, Time 4, and Subject 65, Time 2, were the two that Kohlberg (1973) identified as representing a relativistic, transitional type of thinking that he called "Stage 4 1/2". Note that both of these cases represent sequence anomalies in our analysis as well as the absence of "structured wholeness".

Factor Analyses

Theoretical expectations lead us to believe that moral judgment development is a single general domain cutting across verbal dilemmas and issues. To examine this hypothesis, we factor analyzed the correlations among stage scores on each of the issues across the dilemmas.

Correlation matrices for the whole population and for the 16-17 year age group are given in Tables 15 and 16. The correlations are all positive and moderately high, consistent with our expectations.

Insert Tables 15 and 16

Table 16 indicates that the correlations among moral issues are not due simply to their common correlation with age since the correlations remain positive and moderately high even within a single age group. The 16-18 year age group was used for this analysis because it is the group in which the number of subjects is largest.

Table 17 shows the results of the principal components analysis. The first row of each age group shows the highest and second highest eigenvalue and corresponding percentage of variance accounted for. The succeeding eigenvalues were much smaller and therefore not reported. The next three rows give the highest and second highest eigenvalues for each of the three forms. With all subjects pooled, the second eigenvalue is below one and therefore, by convention, is disregarded. Within the different age groups the second eigenvalue barely exceeds one and adds very little to the percentage of variance contributed by the first factor.

Insert Table 17

Table 18 gives the unrotated first factor loading, for the entire sample on all eighteen issues and for each separate form. The loadings of each issue varied from one dilemma to another, so that no single issue is the best representative of a general moral judgment level.

Insert Table 18

In spite of the marginal contribution of succeeding factors in the age group with the largest number of subjects, we extracted all factors with eigenvalues greater than one. The unrotated loadings for that age group (16-18 years) are given in Table 19. After the first general factor, the loadings on the specific factors are greatly diminished and reveal no consistent patterns that could be interpreted. Similar factor matrices resulted from the other age groupings.

Insert Table 19

In attempting to determine the existence and nature of a multifactoral solution, the factors were rotated using both orthogonal and oblique methods. Here, too, no clear interpretation is evident upon inspection across the five factors. In summary, then, for several age groups and for the sample as a whole, no more than one interpetable factor emerged even when the multiple factors were subjected to orthogonal and oblique rotations. In all cases, the eigenvalue and corresponding proportion of variance accounted for by the first factor far exceeded those of succeeding factors. We conclude, therefore, that moral judgment as measured by the Interview Forms A, B, and C and scored using the Standard Form Scoring Manual, is a single, general domain.

Relation of moral judgment stage to age

As one would expect of a developmental variable, our data show a clear relationship between age and moral judgment stage. The correlation between age and MMS was .78. As shown in Table 20, mean moral maturity score increases monotonically from 188 (Stage 2) at age 10 to 375 (Stage 4(3) at age 36. As shown in Figure 1, the frequency of usage of Stages 1 and 2 decreases from age 10 on, Stage 3 increases up to age 16-18 and then decreases, and Stage 4 begins at 0 and rises monotonically to 62% at age 36. Stage 5 also increases monotonically after its entry at age 20-22 but never rises above 10% of the reasoning in the sample as a whole.

Insert Figure 1

Table 20 presents the data in terms of global as well as moral maturity scores. As that table indicates, the proportion of Stage 1/2's in the sample decreases from a high of 47% at age 10 to 2% (one subject) at age 16 to 18. No subject is below Stage 2 after that. Stage 2's decrease from about one-third of the 10 year olds to 11% of the late adolescents (age 16-18). No one beyond age 18 was scored as Stage 2. Very few subjects at age 20 or beyond were scored at Stage 2/3. Most of our 2/3 subjects were in the 13-14 year age group. The proportion of Stage 3 subjects increases up to age 18, then decreases to zero at our oldest age (36). The transition to Stage 4 does not seem to begin before late adolescence--we found only one Stage 3/4 subject before the 16-18 year age group. The proportion of 3/4's in the sample increases through the early 20's and then levels off to a little less than half the sample. Consolidated Stage 4 did not occur in our sample before age 20 and the proportion of Stage 4's continued to increase through the oldest age in the study (age 36). The transition to Stage 5 occurs even later, with no 4/5's occurring before the mid-twenties. The proportion of subjects reaching Stage 4/5 remains low (11-16%) throughout the age range in which it appears, from the mid-twenties through the mid-thirties.

Looked at from the point of view of age norms in our sample, the interviews of most ten year-olds were scored at Stage 1/2 or 2, a few were scored at 2/3. Most early adolescents (13-14) were 2/3, though some were Stage 2 or 3. Almost half of the late adolescents were scored as Stage 3, about one-quarter had begun the transition to Stage 4, and a little less than one-fifth were still in transition from Stage 2 to Stage 3. Youths in their early twenties were most likely to be intransition between Stages 3 and 4 or to be still solidly at Stage 3. About one-fifth of these subjects had completed the transition to Stage 4. Most subjects from the mid-twenties through the mid-thirties were scored as Stage 3/4 mixtures (or transitionals) with decreasing numbers at Stage 3 and increasing numbers at consolidated Stage 4. Subjects scored as post-conventional (4/5 or 5) represent about one-sixth to one-eighth of the sample from the mid-twenties on.

Curves of mean stage usage for a group do not necessarily represent curves of growth for any of the individuals in the group. Accordingly, we present figures of changes in stage usage over time from four representative individuals. These subjects were randomly chosen from among those who were tested at all six time points and they can be considered typical of our data.

Insert Figures 2, 3, 4, and 5

Subject 3, for example, never shows reasoning at more than two adjacent stages.[6] His reasoning at age 10 is scored at Stages 1 and 2. At age 13, his moral judgment remained dominantly Stage 2 but the Stage 1 has dropped out and has been replaced by substantial usage of Stage 3. At age 14 the Stage 3 has become dominant and Stage 4 has entered as a second stage. This pattern remains stable between ages 17 and 21 but as he reaches young adulthood Stage 4 begins to predominate over Stage 3. Between ages 21 and 25 this trend continues with the proportion of Stage 4 usage increasing and the proportion of Stage 3 decreasing. Perhaps most noteworthy is the orderliness and regularity of the developmental curves with earlier stages dropping out as later stages enter such that the subject seems to be always in transition from one stage to the next. Also noteworthy is the fact that development continues throughout the age range sampled, never reaching a final plateau. The figure for Subject 9 shows a similar pattern with the exception that some use of Stage 1 seems to linger even after the entry of Stage 3 at age 13. (This may be due to the fact that scoring errors confusing Stages 1 and 3 are common when responses are ambiguous or poorly probed.) The rate of development for Subject 9 is slower than that of Subject 3 with Stage 2 still in evidence at age 17 and Stage 4 barely above the error boundary at age 21. Development between ages 25 and 29 is particularly striking in this subject, however, and at age 20 he is scored globally as consolidated Stage 4 while Subject 3 remains globally at Stage 4(3) at that age. Subjects 37 and 42 also show the regular pattern of movement through the stages. These figures are of particular interest in that they show substantial development occurring into young adulthood - between ages 29 and 33. Stage change with age in all four subjects seems to alternate between periods of fairly rapid development and periods of slower development or consolidation. The curves for Subject 3, for example, are especially steep between ages 13 and 17, plateau to some degree between 17 and 21, and become steep again indicating a second growth period between ages 21 and 25.

Furthermore, in all four cases a fairly saltatory model of development appears justified. That is, the absence at each time point of three stage mixture not only supports the assumption that reasoning at a given time forms a structured whole, it also bears on the process of development. The point here is that for the most part Stage 1 has dropped out by the time use of Stage 3 has begun, Stage 2 has dropped out by the time use of Stage 4 has begun, and so on. The holistic character of the developmental change is evident across the entire age range in our study.

Cohort Effect

Our sample can be considered to be composed of subjects from three cohorts - those who were 10 years old at Time 1, those who were 13 at Time 1, and those who were 16 at Time 1. When data from these three groups are analyzed separately, we find that they all show the same moral judgment patterns. That is, subjects from all three groups demonstrate "structured wholeness" and invariant stage sequence in their moral judgment. As is shown in Figure 6, the mean moral maturity scores for the three groups differed somewhat, particularly in that the means for the Time 1 sixteen-year-olds were lower at each testing time than were the means for the Time 1 ten-year-olds. The difference is not significant, however, and is most likely due to sampling procedures. The ten and thirteen year olds were chosen from an elementary school while the sixteen year olds were chosen from a high school in the same community. Although each cohort was composed of half working class and half middle class subjects, there was a tendency for the high school subsample to come from somewhat lower socioeconomic origins within the broader working class and middle class categories than was true for the younger subjects. That is, the high school tended to draw the lower end of both the working class ("lower" rather than "upper lower" class) and the middle class (lower middle and middle rather than upper middle class) categories. This may have resulted in the consistent though non-significant tendency for the Time 1 sixteen years olds to score a little lower than the others at each testing time. Based on the information available to us, there is little reason to believe that the slight group mean differences represent a historical cohort effect. Neither do we interpret the data as indicating a practice effect on the moral judgment interview since in our test-retest analysis we found no such practice effect.

Stability of Individual Differences

Our data bear on the question of whether there are stable individual differences in rate of moral judgment development. The issue here is whether scores at one age predict to scores at a later age. As shown in Table 21, the correlations between moral maturity scores at age ten and scores at ages 13, 17, and 21 are positive and significant but not very high. The age 10 scores do not predict adulthood scores at all. In contrast, MJ scores at age 13 are much more highly correlated with later scores, even with scores at age 32-33. Whereas the highest correlation between age 10 scores and those from a later testing was .47, the lowest correlation between age 13 scores and later ones was .46 and the correlations ranged up to .70. Age 16 scores do not seem to predict later scores any better than do the age 13 scores, but at age 20 the scores again seem to increase in stability to some degree.[7] It seems, then, that there are two periods during which stabilization occurs. The first occurs between ages 10 and 13, that is, as the subjects enter adolescence. The second period of stabilization occurs between the ages of about 17 and 21, that is, as the subjects enter early adulthood. This does not necessarily imply that these are periods of particularly rapid development in moral judgment. In fact, the mean increase for the group as a whole was not much greater between ages 10 and 13 than between ages 13-14 and 16-18, nor was it greater between 16-18 and 20-22 than between 13-14 and 16-18. As was shown in Table 20, the increase in mean MMS with age is gradual and quite even across the age range studied. Rather, these periods of stabilization reflect a tendency for subjects to shift their positions in relation to each others' moral judgment scores as they enter adolescence and again as they enter adulthood.

Insert Table 21

The Relationship of Moral Judgment Stage to Other Variables

Given the existence of fairly stable individual differences in moral judgment stage, the question arises as to what determines these differences. Our study provides data on a number of variables that might be expected to be related to moral judgment stage. These variables include socioeconomic status of origin, sociometric status (as measured at Time 1), IQ (group test scores taken from school records at Time 1), and educational level attained.

Socioeconomic status. As Table 22 indicates, correlations between parents' SES and subjects' moral judgment scores were moderate at every age. They ranged from .32 to .62 and all but one were above .40. There was no tendency for the relationship to either strengthen or attenuate with age and the fluctuations of correlations across age showed no clear pattern.

Insert Table 22

Perhaps more enlightening is a comparison of working class and middle class subsamples on mean percentage of each stage used at each age. As shown in

Figure 7, the distributions are quite different for the two subsamples. Both Stages 3 and 4 exceed the 10% error cut-off level at earlier ages in the middle class group. Stage 3 is present at age 10 for middle class subjects, but not until age 13 for working class subjects. Stage 4 appears at age 16 in the middle class group, but not until age 20 in the lower SES group. Furthermore, Stage 2 remains above the 10% cutoff somewhat longer for lower SES subjects. Whereas Stage 5 usage exceeds 10% at age 28 in the upper SES group, it never exceeds the error cut-off in the lower SES group.[8]

Insert Figures 7a and 7b

Sociometric status. Correlations between moral maturity score and sociometric status were substantially lower than correlations between moral and judgment scores and SES. Except at age 10 they did not achieve significance. Again no clear pattern emerged in fluctuation of SMS/MJ correlations across time. (See Table 22).

A comparison of stage x age figures (Fig. 8a and 8b) for sociometric integrates and isolates reveals that Stage 3 exceeds the 10% cut-off at age 10 in the integrated group but does not do so among the isolates until age 13. In addition, the isolates at age 10 show substantially more Stage 1 use than do 10 year old sociometric integrates. Stage 1 drops below 10% at age 13 for integrates but not until age 16 for isolates. As is reflected in the correlations, differences in sociometric groups beyond age 13 are mininal. Overall, then, the data show that at the time that sociometric status was evaluated, integrates tended to show more Stage 3 and less Stage 1 than sociometric isolates. However, the difference between groups is not maintained over time.

Insert Figures 8a and 8b

A multiple regression analysis indicates that age accounts for 60% of the variance in moral maturity scores. Adding SES raises the proportion of variance accounted for to 67%, and adding sociometric status adds very little, raising the proportion of variance accounted for to 68%. Since SES was not reassigned for adult subjects who may have changed in social class and since sociometric status was assessed only at Time 1, these figures on the relative importance of age, SES, and SMS should be interpreted cautiously.

Intelligence. Correlations between moral maturity score and IQ were non-significant and ranged from .17 to .27 in childhood and adolescence for our sample but become substantially higher at age 24 and above (.37-.59). (See Table 22.) Except at age 36 (an analysis which included only ten subjects), the correlations increase steadily from age 24 on and are significant at the three age levels between 24 and 33. This occurs in spite of the fact that IQ was assessed only at Time 1. It appears, then, that while rate of moral development in childhood and adolescence is only slightly related to IQ, the final level achieved in adulthood is more closely related to intellectual capacity, perhaps partly via differential educational experiences that are related to intelligence.

Educational level attained. Correlations between adult moral maturity score and educational level attained ranged from .54 to .69, (p .05) with no clear relationship between strength of correlation and age. (See Table 22.) (Only ages 23, 27 and 31 are reported since formal education for many subjects was not completed before age 23.) Note that these correlations are somewhat higher than the correlations between moral judgment and either social class or IQ.

Partial correlations substantiate the interpretation that moral judgment stage is related to educational experience itself rather than to educational level as a reflection of IQ and SES differences. The correlation between MMS and subject's education with IQ partialled out is .36 (p .05) at age 28, the only age at which the N was large enough to make a partial correlation feasible. The correlation between MMS and subject's education with parents' SES partialled out is slightly higher, r=.45 (p .05) at age 28. The correlation between MMS and education with both IQ and SES partialled out was somewhat lower but still significant, r=.26, p .05. It appears, then, that educational experience per se is related to moral judgment maturity beyond the association of education to SES and IQ.

Table 23 presents the moral judgment x education data in a way that is perhaps more revealing than a correlational analysis. The table indicates that no subjects from either working class or middle class origins achieved consolidated Stage 4 moral judgment without attending at least some college and no subject from either social class achieved Stage 4/5 without having completed college. (Recall in this regard that no one reached Stage 4 before age 20 and no one reached Stage 4/5 before age 24.) Whereas more middle class than working class subjects reached Stages 4 and 4/5, both of these moral judgment levels were attainable by subjects of working class origins who had attended college. (No statistical analysis was performed on these data since expected values in the tables' cells were less than 5.)

Insert Table 23

Chapter V

DISCUSSION

Standard Scoring Reliability and Validity

Taken together, the data presented here indicate that Standard Issue Scoring has succeeded in addressing the central methodological concerns raised in critiques such as the widely cited paper by Kurtines and Grief (1974). The criticisms of Kohlberg's method presented by Kurtines and Grief fall into three general categories: (1) Non-standardization of the interview and coding scheme; (2) questionable reliability of the coding schemes, expecially important given their complexity and "subjectivity"; and (3) questionable validity of the coding schemes, particularly a failure to present clear evidence of invariant stage sequence.

As its name indicates, The Standard Issue Moral Judgment Interview and Scoring System represent an attempt to standardize the assessment of moral judgment stage. The nine hypothetical dilemmas constitute three parallel interview forms, the probing questions are specified, and the scoring system is no longer undergoing revision.[9] Although the system remains complex, it can be mastered by individuals who are relatively inexperienced as indicated by the inter-rater reliability figures presented here. Test-retest and alternate form reliability are also high as is internal consistency of scores assigned an interview. This should not only assure reliable scoring within each study that uses the measure, but should also provide enough standardization to allow comparison of results across studies by different investigators. This increase in reliability and objectivity in scoring was achieved by the creation of a small but conceptually, rather than arbitrarily, defined unit of analysis, by the elaboration of clear rules for classification of material and for defining a match betwen manual and interview judgements, and by the construction of specific and detailed stage criteria, the critical indicators, which specify exactly what is required in the interview in order for a stage score to be assigned. On the other side of the reliability/validity balance, structural explications of each manual item were provided in order to minimize loss of validity due to overly literal interpretation of the manual.

With regard to validity, we have argued that prediction to an external criterion such as action taken in a moral conflict situation is not an appropriate indicator of the instrument's validity. The appropriate question is whether the interview and scoring system provide a valid assessment of moral judgment stage, not of moral character as a whole. If validity is understood in this way, the present longitudinal study can be considered to provide substantial support for the validity of Standard Issue Scoring. That is, when Standard Issue Scoring was used to score longitudinal interviews not used to construct the measure, it yielded scores that agreed very closely with the theoretical predictions of invariant sequence and internal consistency.

In order to fully establish the validity of Standard Issue Scoring, there is a need for further research on discriminant and construct validity, particularly longitudinal studies of populations differing from that described here. Four such longitudinal studies using Standard Issue Scoring have been reported: one with American girls and young women (Erickson, 1980), one with American male and female college students (Gilligan and Murphy, 1979), one with Turkish boys and young men (Nisan and Kohlberg, in press), and one with Israeli boys and girls (Snarey, 1982).

The studies conducted in the United States are difficult to interpret because of design problems. Both Erickson's and Gilligan and Murphy's samples participated between Times 1 and 2 in courses intended to promote moral development. Particularly problematic is the fact that these courses included direct study and discussion of Kohlberg's theory. Furthermore, these studies used different forms of the moral judgment interview at different testing times rather than maintaining a consistent set of dilemmas across all waves of data collection. Since scores on different dilemmas may differ somewhat, the unsystematic variation in dilemmas used may introduce error into the data. A further limitation of these studies is the fact that Erickson's subjects at each testing time and Gilligan and Murphy's at Time 1 and Time 2 filled out a written version of the moral judgement interview rather than participating in individually administered taped interviews. Since responses to written interviews are less well elaborated than responses to oral interviews, mis-scoring of the former is more likely to occur.

Possibly as a result of these methodological problems, both Erickson and Gilligan and Murphy report somewhat more frequent sequence reversals than were found in the present study. Over three three to eight year intervals, the Gilligan and Murphy data show 12-15% downward stage movement as compared to our 4%. Over a four year interval (between T1 and T5), the Erickson data show 9% downward stage movement. Calculated over all one year testing intervals, Erickson finds 13% downward stage movement. It should be noted that in spite of indicating somewhat more frequent stage reversal, all of these figures fall within the limits of measurement error as represented by our test-retest data.

The difference between Erickson's one-year and four-year interval data is consistent with earlier findings (Kuhn, 1976; White, Bushnell, and

Regnemer, 1978) that the longer the time interval between subsequent testings, the more likely a subject was to have advanced rather than regressed. These results have been interpreted as evidence that short-term fluctuations in moral judgement occur in the context of long-term progression. Whether it is most appropriate to treat the short-term fluctuations as measurement error or as actual stage change is impossible to determine with certainty. Of course, the present study does not allow us to compare short-term fluctuation with long-term progression within the same individuals since all interviews were conducted three to four years apart. However, we have seen that downward changes from T_1 to T_2 in our test-retest sample are substantially more frequent than from Tn to Tn+1 in our longitudinal sample.

We interpret the differences between T_1 and T_2 in the test-retest sample as measurement error rather than as true developmental fluctuation, primarily because the percent agreement and correlation coefficients are essentially the same for test-retest, inter-rater and alternate form reliability. That is, there seems to be slippage or error of about 1/3 stage in the system. When this confidence interval is superimposed on negligible developmental progress as in the case of testings one year apart, the result will be relatively more frequent occurrence of lower scores at Tn+1 than at Tn. When the measurement error is superimposed on developmental progress as in the case of testings three years apart, the occurrence of a score 1/3 stage below the individual's "true stage" would result in either no change from Tn to Tn+1 or a smaller positive change than would otherwise appear. If most individuals move forward at least 1/3 stage in four years, this interval will not yield many cases of apparent regression.

The two cross-cultural longitudinal studies that used Standard Issue Scoring more closely parallel the present study in design. Subjects did not participate in programs of moral education, dilemmas used were held constant across all waves of data collection, and interviews were individually administered. In both these studies, the stage sequence data were as clear-cut as in the present American longitudinal study. Nisan and Kohlberg (in press) report the results of data collected over a twelve year period in Turkey. Subjects were 23 boys ranging in age from 10 to 17 at Time1. Interviewing took place in 1964, 1966, 1970, and 1976 in three locations in Turkey: a rural village, a seaport provincial capital, and the national capital. Stages one through four were present in this sample. There was no stage skipping and only one instance (3%) of downward stage movement. Snarey (1982) reports 10 year longitudinal data on 92 boys and girls from Israeli cities and Kibbutzim. Snarey found that the Standard Issue Scoring System yielded reliability data on Israeli interviews comparable to the reliability data reported here for our American sample. Stages 2 through 4/5 were present in this sample. With regard to invariant sequence, he found that downward stage change occurred in only 5% of the adjacent time points.

To sum up, there have been four reported longitudinal studies of moral judgment using Standard Issue Scoring in addition to the present study. In two of these studies there were methodological problems which might be expected to increase measurement error. The stage sequence data are in factsomewhat less orderly in these two studies than in the others. However, the frequency of sequence anomalies even in these studies is quite low, substantially lower than the frequency of downward stage change in our test-retest data. In the two cross-cultural studies the sequence data are essentially identical to those reported in the present study. Of these four studies, only one, Snarey, presents internal consistency data. The results of Snarey's Israeli study regarding internal consistency of moral judgment are essentially identical to our own results. That is, in a factor analysis only one general factor emerged, and very few subjects showed reasoning at more than two adjacent stages.

In light of criticisms of Kohlberg's theory by Holstein (1976), Gilligan (1977), and others, one might wonder whether subjects in the studies that report more orderly sequence data are more predominantly male and subjects in the studies with less orderly data are more predominantly female. The fact that Erickson's subjects were all female and Nisan and Kohlberg's subjects as well as those in the present study were all male may seem to support this intepretation. However, the subjects in Gilligan and Murphy's and Snarey's studies were approximately half male and half female. The issue of sex differences in orderliness of stage sequence may be addressed by calculating separately the percentage of reversals among males and females in the Gilligan and Murphy and Snarey studies. In fact, there do not appear to be significant sex differences in either of these studies in either developmental stage or in orderliness of stage sequence.

Overall, these four studies using Standard Issue Scoring are consistent in yielding more orderly sequence data than have been reported by investigators using earlier scoring systems. We have already referred to Kramer's analysis of Kohlberg's data and we shall discuss it further in the next section.

Holstein (1976) has reported three-year longitudinal data on Kohlberg's moral judgment interview with 52 thirteen-year old boys and girls and their parents. Using Structural Issue Scoring on interviews that were primarily written, she found that a substantial number of subjects skipped stages or reverted from a higher to a lower stage between T_1 and T_2. Holstein found both age and sex differences in the frequency of stage skipping. While 21% of the adolescents skipped at least one stage between T_1 and T_2, only 7%

of the adults did so. Stage skipping in males tended to be from Stage 1 or 2 to Stage 4, in females from Stage 3 to Stage 5. Holstein also reported a substantial amount of regression from T_1 to T_2 along with a dramatic difference in frequency of regression for lower stage (1-3) subjects and for higher stage (4-6) subjects. While virtually none of the lower stage subjects regressed, many (20-33%) of the higher stage subjects did revert to a lower stage at T_2. Without rescoring Holstein's interviews, it is impossible to determine whether the differences between her results and those presented here are due to scoring differences, sampling differences, or to some other feature that differs across the two studies. Holstein interprets the high rate of regression among higher stage subjects as an indication that Kohlberg's stages 4 through 6 do not represent an invariant developmental sequence but rather are alternative and equally mature forms of moral judgment. An alternative interpretation is the higher stages were inadequately defined in scoring systems prior to Standard Issue Scoring. Perhaps the strongest evidence for the validity of this interpretation is the fact that Kramer's analysis of Kohlberg's longitudinal data showed paterns of regression and stage skipping that were almost identical to those Holstein found. Upon rescoring using Standard Issue Scoring, virtually all of the anomalies in Kohlberg's data disappeared. That is, in the study reported here, Stage 4 and 5 do appear to be developmentally ordered. The relation between Stage 5 and 6 cannot be addressed, of course, since Stage 6 has been dropped from the system. Among the other longitudinal studies that used Standard Issue Scoring, only two traced the stage sequence up to Stage 5. The results with regard to sequence in these two studies are inconclusive. While Snarey (1982), finds virtually no downward stage change at either low stage or high stage levels, Gilligan and Murphy (1979) do report a tendency for subjects to regress from Stage 5 to transitional level 4/5. This difference may be an artifact resulting from the fact that the latter subjects had studied Kohlberg's theory while the former subjects had not. On the other hand, it might indicate some remaining unresolved inadequacy in the differentiation of Stage 5 from transitional level 4/5.

White, Bushnell, and Regnemer (1978) present longitudinal data collected on the island Eleuthera in the Bahamas. Their data include a 3-7 year longitudinal sample and a series of 2-7 year longitudinal samples. Like Holstein, they used Structural Issue Scoring. Although the findings indicate a general pattern of upward stage change over time, some respondents in all subsamples did regress. Since the authors report analyses of variance rather than frequency of regression, we cannot directly compare their results with those from the other studies discussed here. White et al point out that in their data, the longer the time interval between subsequent testings, the more likely a subject was to have advanced rather than regressed. They argue that their findings suggest that moral judgment stages develop sequentially, but short term fluctuations either up or down may often take place in the context of long term progression. This is the same pattern that Erickson (1980) found using Standard Issue Scoring. We have argued above that the pattern is probably an indication of measurement error superimposed on steady developmental progression.

Another instance of this pattern has been reported by Kuhn (1976) in a one year longitudinal study of 50 five to eight year old children. Kuhn used simplified versions of the dilemmas from Standard Form A and used global story scoring to analyze the responses. The children were interviewed three times - at Time 1, 6 months after T_1, and one year after T_1. Kuhn found that at six month intervals the data showed a considerable amount of downward as well as upward stage movement. In fact between T_1, and T_2 and between T_2 and T_3, subjects were not significantly more likely to move upward than downward. The pattern of change from T_1 to T_3 was drastically different, however. While there was a substantial amount of upward stage change, only one subject out of fifty (27%) showed any downward movement (using our nine point scale). There were no occasions of stage skipping. It is worth noting that at this age level one year constitutes a long enough interval to show steady progression while at older ages one year intervals show the kind of short-term fluctuation that Kuhn found in her six-month intervals. This is no doubt due to age differences in rate of development.

Also noteworthy is the fact that the sequence in Kuhn's one year data is as orderly as it is in our data in spite of the fact that she used an earlier scoring system. This further illustrates the point made earlier that the higher stages were more radically redefined in Standard Issue Scoring than were the lower stages.

Turning to our own longitudinal data, we see a dramatic improvement over Kohlberg and Kramer's earlier analysis in the fit between theory and data in relation to both structured wholeness and invariant sequence. Before considering the details of that improvement, we must raise again the question of circularity. Is the improvement simply the result of post hoc manipulations that ensure our obtaining the desired results, i.e., it couldn't have been otherwise? We think not. As indicated earlier, Standard Issue Scoring rules do not force consistency of stage scores assigned an interview. Scores could have been scattered across three or four stages for an indicidual subject, yet they were not. Our design prevented more subtle halo effects by requiring that each of the nine dilemmas for each subject be scored independently of one another. Furthermore, the interviews from which sequence and internal consistency data were derived were not used at all in reformulating the stages or creating the scoring manual. If there were no clear sequence in moral judgment development, it should not have been possible to derive a system from seven "construction cases" that would yield such clear sequence for the 51 blind cases. The fact that the raters were required to use specific, objective criteria in assigning scores makes the sequence and internal consistency findings even more convincing. The records kept on each subject document in exactly what sense he was judged to have developed (or remained the same) from one time to the next and in exactly what sense his responses can be considered structurally consistent from one dilemma to the next at the same testing time.

Let us look, then, at the improvement of theory/data fit by comparing our results with those of Kramer (1968) using the early scoring systems of Sentence and Story Rating. Figure 12 presents the average distribution of stage use for Kramer's analysis and our own. There is clearly a much greater spread of scores for Kramer's subjects across non-adjacent stages. While Kramer reports an average of 16% of a subject's scores are assigned at stages two steps removed from his modal stage, in our sample only 1% of scores were assigned to those stages. Whereas none of our subjects used reasoning three steps removed from the modal stage, Kramer did find 5% of such use. (Recall that according to our scoring rules such an occurrence is perfectly permissible but nevertheless, empirically it did not occur.) We also see an improvement in the sequence data. While Kramer's scores show 19% downward stage change (using the 9 point scale), we find only 4%.

What accounts for this improvement? One factor is surely the increased reliability of Standard Issue Scoring. If scores are unstable due to unreliability, there will be greater heterogeneity across dilemmas and more frequent sequence anomalies due to measurement error. But we have argued that Standard Scoring is more valid then earlier systems as well as more reliable in that it more adequately differentiates moral judgment content from the organization or structure of moral judgment. If content associated with structure is used as a scoring criterion, the data may yield appropriate trends, but exact sequence will not be found, as was true in Kramer's scores. Because content is only probabilistically related to structure, a subject may begin to use the content associated with a higher stage before he has used the content usually associated with a lower stage. Whereas a structurally accurate evaluation might show an increase in the sophistication of this individual's reasoning, a content analysis would show regression. Also, in terms of internal consistency, if scores are not assigned on the basis of the truly developmental features of the judgments, then those scores may or may not be equivalent across dilemmas or issues.

Theoretical Evolution

The method we have used in responding to early theory/data discrepancies is not strictly hypothetico-deductive. In other discussions we have called this method "bootstrapping" (Colby, 1978; Kohlberg, 1980; see also Loevinger, 1976). By "bootstrapping" we are referring to the notion of an evolving research program in which data feed back into theory and method to improve the account of development we offer. This process is not unique to our efforts. In fact, it could be argued that it is the model used in all scientific endeavors.

In justifying the legitimacy of this model, we shall consider and respond to a critique of Kohlberg's theory by Nicolayev and Phillips (1979). (See Puka, 1979, for a more complete reply to Nicolayev and Phillips.) These authors have attempted to evaluate Kohlberg's theory in relation to criteria set forth by Lakatos (1976). As Nicolayev and Phillips point out, Lakatos distinguishes between a theory's "hard core" assumptions (those elements of theory that cannot be changed without abandoning the theory) and its "protective belt" (those aspects of the theory that are subject to change in light of experience). A progressive research program includes "a partially articulated set of suggestions or hints on how to change, develop the 'refutable variants' of the research programme, how to modify, sophisticate, the 'refutable' protective belt." The resulting modification of the "protective belt" must anticipate new facts or be "content-increasing" rather than simply be means to explain away discrepant findings.

While we agree with Nicolayev and Phillips in identifying the structural wholeness and invariant sequence of stages as defining the hard core[10], we are led to conclusions very different from theirs. Our point here is not to argue that the data support these assumptions (as they clearly do), but rather to defend the theoretical revision process as progressive rather than "a readjustment of scoring procedures to give the results required by the theoretical assumptions." We have already pointed out that the procedure for revising the stage descriptions, constructing the manual, and scoring the data did not in any sense guarantee the results that were obtained. The theoretical and methodological evolution did not consist in generating new scoring rules that would force internal consistency or postulation of new stages to explain sequence regression. In fact, the two interviews identified by Kohlberg (1973) as "Stage 4 1/2" are still counted as violations of sequence in the current analysis. Instead the evolution entailed a sweeping revision of the entire stage sequence and a radical redefinition of the basic structures in moral judgment development. These revisions have not only permitted the generation of new and testable hypotheses (for example, the predictive value of moral judgment substage for moral conduct, see Kohlberg and Candee. 1981). They also constitute theoretical advances in ways that go beyond the fact that they yield improved reliability, internal consistency, and sequence data. One example is of particular interest in light of Phillips' and Nicolayev's argument that there is no "logical necessity" to the order of the moral judgment stages. Although we shall not take up here the critics' confusion as to what Kohlberg means by logical necessity (see Puka for this), we can provide an illustration of the sense in which the current stage descriptions allow a clearer rationale for the invariance of the sequence than did the early, more content-based stage descriptions. If Stage 3 is considered the "interpersonal concordance orientation" and Stage 4 is defined as the "law and order orientation", there is no discernible reason for Stage 4 to develop only after one has passed through Stage 3. Neither is there any clear justification for the greater adequacy of law and order thinking. In fact, to most liberal social scientists this claim must seem counter-intuitive, even objectionable. Fortunately, the more formal stage criteria of the current system provide a solution to this problem. Development in level of perspective is one important contributor to the increasing cognitive adequacy of each stage. To oversimplify a bit, development through the stages yields increasingly better understanding of and ability to intergrate several points of view on a moral conflict situation along with a greater ability to distinguish morally relevant situational factors and to take them into account. With regard to the question of each stage presupposing the understanding achieved at lower stages, the level of perspective notion again provides a convincing rationale. The claim is simply that one cannot take a prior-to-society perspective without the capacity to think in terms of a social system; one cannot take a social system perspective without the capacity to think in terms of shared norms and expectations of a group or mutually oriented dyadic realtionships; one cannot respond in terms of coordinated individual perspectives until one can differentiate individual perspectives, and so on. In addition to this very general sense in which each stage presupposes development through the previous tages, Standard Scoring clarifies the increasing sophistication, differentiation, and integration of specific moral concepts such as promise-keeping or the importance of intentionality in assigning punishment. Consider, for example, the following items from the Form A manual:

SHOULD THE JUDGE PUNISH HEINZ IF HE DOES STEAL THE DRUG?

Stage 2 - The judge should be lenient because Heinz wanted to keep his wife alive.

Stage 3 - The judge should be lenient because Heinz acted unselfishly in stealing the drug for his wife.

Stage 4 - The judge should be lenient because (s)he should recognize the extenuating or mitigating circumstances and be lenient within the parameters of the law.

Stage 4/5 - The judge should be lenient because (s)he can find a precedent or develop a rule of law that reflects what is right.

These examples illustrate a developmental progression in the understanding of intentionality in relation to punishment. At Stage 2 the argument for leniency is that Heinz had a good reason for stealing the drug. That is, it is assumed that there is no need to punish transgressions which make pragmatic sense. At Stage 3, the argument expands to include a concern for good motives. It is no longer enough for the reason to be pragmatically sensible. At Stage 3 the intention must be prosocial and its significance derives from its implications for the actor's underlying good character. That is, it is assumed that one ought not to punish someone who is good, well-intentioned, unselfish, and so on. Although the individual who reasons at Stage 4 appreciates that Heinz had a good reason to steal and was well-intentioned, having good intentions is no longer seen as sufficient to justify leniency toward a lawbreaker. The Stage 4 judgment reflects an understanding of a systematic, formalized way to take intention into account within the established legal framework. That is, leniency is justified on the basis of a judge's discretion within a framework that assumes the value of maintaining legal consistency. The level 4/5 judgement goes beyond the Stage 4 conception of maintaining legal precedent toward a recognition of the legitimacy of creating precedent through interpretation of the law. It should be apparent that each stage involves an awareness of the considerations of previous stages and some recognition of the legitimacy of those considerations. As one enters a new stage, however, the argument from the previous stage is no longer seen as sufficient. At the new stage the argument is qualified and placed within a more complex framework which transforms its meaning.

Age Norms

An important aspect of the change in moral judgment scoring criteria is that the current system stresses more heavily the need for the rater to look beyond the subject's superficial verbal responses to their conceptual significance for the individual. This is achieved methodologically through intensive probing of statements made in the interviews (What do you mean by that? Why is that important? and so on) and by requiring each interview/manual match to undergo a structural evaluation based on the stage structure paragraph provided with each criterion judgment. One result of this shift is that people are less likely to be "given credit" for clichés or language which resemble that of higher stages if they do not exhibit the appropirate conceptual underpinning. This means that stage criteria are in general more stringent than in earlier scoring systems. The result is a fairly radical change in age norms.

Figure 10 presents percentage of stage usage for the middle class longitudinal subjects aged 10 through 24 using Sentence Rating scores (10a) and Standard Issue scores (10b).[11] The distributions are clearly very different. One noteworthy difference is that the Sentence Rating scores show spread across a wider range of stages for each age. This no doubt reflects a tendency for the early scoring system to treat content differences as developmentally significant both within and across subjects.

Along with the greater variance in the Sentence Rating scores, we see a tendency for Sentence Rating scores to be higher than those using the Standard Issue System. In the early analysis, Stage 4 is in evidence at age 10 and is relatively heavily used at ages 13 and 16. In our analysis there is no use of Stage 4 at age 10, less than 10% at 13, and less than 20% use of Stage 4 even at age 16. The difference in use of Stage 5 is even more dramatic. Sentence Rating assigns some Stage 5 scores to subjects as young as age 13 and shows Stage 5 as the group's most frequently used stage at age 16. We show no use of Stage 5 until age 24.

As subjects move from age 16 to 20, Stages 4 and 5 reverse in the early analysis. This is, at age 16 Stage 5 predominates over Stage 4 while at age 20 Stage 4 predominates over Stage 5. This reversal indicates the presence of sequence anomalies in the early data such that a number of subjects scored at Stage 5 at age 16 were scored as Stage 4 at age 20. The Standard Issue data show no such reversal. Between ages 16 and 20 Stage 4 usage increases steadily and Stage 3 usage decreases. There is no evidence of Stage 5 until age 24 and Stage 5 usage does not decrease once it has begun to occur. This results in a very different picture of stage usage in late adolescence and early adulthood than had been assumed previously. Accoring to Standard Issue Scoring, moral judgment at age 20 is essentially completely conventional, whereas using the 1958 system, subjects show substantial representations of Stages 4, 5, 3, and 2.

In summary, both within and across subjects, Standard Issue scores show greater homogeneity than scores reported in earlier analyses of these data. The range of moral judgment stages used by middle class American boys at a a given age seems to be fairly limited. From our current perspective it is clear that Sentence Rating profiles were reflecting a range of content emphases or orientations rather than a range of qualitatively different structures. Superficial differences which in the earlier system were treated as structurally significant are no longer represented by different stage scores.

One important issue raised by the difference in the two analyses is the question of adulthood stages. In light of the emergence of Stages 5 and 6 in adolescence, Kohlberg and Kramer (1969) concluded that "there was no new way of thinking about the moral situtations that was found in adulthood and not found in adolescence" (p. 105).[12] Our interpretation of this finding is that the early scoring systems could not discriminate properly among superfically similar moral judgments at Stages 3, 4, and 5. This meant that when "true Stage 5" began to emerge in early adulthood, it was not recognized as a qualitatively new form of reasoning. We can now discriminate principled reasoning from superficially similar conventional reasoning and we no longer see Stage 5 being used in junior high or high school. This means that there _is_ a new stage in adulthood and that many of our subjects continued to develop in their 20's and 30's, rather than reaching a ceiling in mid-adolescence. This result is intuitively appealing since thirty-year-olds often do seem to be more sophisticated in their moral judgment than do sixteen-year-olds. However, it does raise the question of whether Stage 5 is a "naturally developing" stage in the strict Piagetian sense. (See Gibbs, 1978).

Along with the age shift in onset of Stage 5, we also see a reduction in the prevalence of principled moral judgment once it begins to appear. In our analysis, Stage 5 scores do not rise above 15% even at age 36 whereas Kramer shows almost 30% Stage 5 usage at age 16. According to Standard Issue Scoring, the percentage of individuals who reach at least the 4/5 transition in adulthood is only 16%. Kramer reports that about 30% of the subjects in his analysis have at some point shown substantial usage of Stage 5.

Individual Differences

Our results support the assumption that relative maturity of moral judgment is a fairly enduring characteristic of individuals throughout their development and becomes even more stable in adulthood. That is, if an adolescent is advanced in his moral judgment relative to his age peers, he is likely to be advanced in adulthood as well. However, the correlations are not so high as to preclude some changes in relative position for many subjects. Bloom (1964) has suggested that shifts in stability of individual differences in intelligence are useful in defining open periods for educational intervention. There are a number of problems in taking this idea literally in the domain of intelligence (Kohlberg, 1967) and in moral judgment. However, the fairly dramatic increase in correlation with adult scores that occurs between the ages of ten ($r = .20$) and thirteen ($r = .70$) is suggestive that the period before 13 years may be a particularly fruitful one for educational stimulation or intervention. This implication of the findings is discussed as length elsewhere (Kohlberg, 1970).

Social Role and Status Variables and Moral Development

The variable of sociometric status was included in the original design in order to illuminate Piaget's (1965) argument that peer group participation is an important determinant of moral judgment maturity. Our data provide some very limited support for this hypothesis in that our sociometric integrates used more Stage 3 and less Stage 1 than the sociometric isolates. Of course, there is no way of knowing whether these results are due to Stage 3 children being more popular or whether more active peer group involvement and social interaction facilitate the transitions out of Stage 1 and into Stage 3. Both are theoretically possible and, in fact, both may have occurred. In any case, the effect of sociometric status in childhood, at least as measured here, does not seem to extend beyond the transition to Stage 3.

Differences in socioeconomic status, on the other hand, do seem to be associated with moral judgement differences across the entire stage range. Whereas peer group participation may be especially important for the onset of Stage 3, social class seems to be related to the development of Stages 4 and 5 as well as 3. Many more middle class than working class subjects reached both consolidated Stage 4 and the post-conventional 4/5 level, although Stages 4 and 4/5 do seem to be accessible to college educated subjects from working class origins. In previous writing, Kohlberg (1969) has interpreted social class differences in rate and terminus of development as reflecting differential participation in and identification with the society and its secondary institutions. It is argued that this differential participation creates differential role-taking opportunities for middle class and working class children which allow those from the middle class to experience, for example, being integral participants in the society and thus to develop the social system perspective that characterizes Stage 4. Since the design controlled for IQ, the relation between moral judgment and SES cannot be accounted for by IQ differences between working class and middle class subjects.

Educational attainment is, however, confounded with social class. Formal education is itself a very important experience that is more often available to middle class individuals and may mediate at least to some degree the relation between moral maturity and social class. Whereas only 37% of our working class sample completed college, 71% of our middle class sample did so. However, when the variables of social class and educational attainment are separated as in Table 23, it appears that social class is related to moral judgment stage even with educational level held constant. Although the small n's preclude any definite conclusions, the table indicates that, except for the "high school only" group, the scores are higher for middle class than working class subjects within each of the three education levels - some college, completed college, and graduate school. This suggests that the differential experiences that middle class and working class children undergo include but go beyond differential opportunities for formal education.

Formal Education and Moral Development

The finding of a moderate relationship (r = .54) between moral judgment stage and education is not surprising, given the cognitive-developmental nature of the moral judgment stages. As noted earlier, we found that none of our subjects reached Stage 4 without having attended some college, and none reached the 4/5 level without having completed college. This does not imply that college study is always necessary for development of consolidated Stage 4 or movement to Stage 5. In an era or a culture with less emphasis on formal education or less accessibility to college, one might find many self-educated people at Stage 4 or 5. However, the relationship does suggest that development to the higher stages is facilitated by educational experience. Using Rest's Defining Issues Test, G. Rest (1977) found a comparable moderate correlation (r = .45) between education and moral judgment in adults.

IQ and Moral Development

The cognitive basis of Kohlberg's stages is also reflected in moderate correlations between IQ and moral judgment. Although the positive correlations point to the cognitive component in moral development, the modest size of the correlations (.37 - .59 for our sample) indicates that moral judgment is not reducible to intelligence. As in the case of educational experience, correlations between IQ and moral judgment as measured by the DIT are comparable to those between IQ and moral judgment in our sample. Rest (1979) reports that most studies have found IQ x DIT correlations in the .20's to .50's range. (Recall that ours range from .37 to .59 between age 19 and 36.)

College Relativism and "Regression" in the Longitudinal Sample

Readers familiar with the work that was generated in response to Kramer's finding of sequence anomalies may wonder what happened to "Stage 4 1/2". This phenomenon--an apparent regression from Stage 5 to Stage 2 or 3 in late adolescence--was first interpreted (Kohlberg and Kramer, 1969) as a "functional regression in the service of the ego". Kohlberg (1973) later reinterpreted the phenomenon as a structural progression characterizing the transition from conventional to principled moral judgment for some subjects.

In fact, only one of the anomalous cases in the original analysis can be attributed to the phenomenon Kohlberg called Stage 4 1/2. As Table 13 indicates, Standard Issue scores for Case 2 went from a 4(3) score at the end of high school to college period scores of Stage 3 on Form A, Stage 2(3) on Form B, and Stage 3(4) on Form C. This scatter and regression in moral judgment scores coincided with non-scorable metaethical views of ethical relativism and ethical egoism. On follow-up four years later, Case 2 had moved to a postconventional score of 5(4) as Table 13 indicates. Case 2 was one of the cases discussed in the Kohlberg (1973) account of college transitional relativism.

A second case discussed as "Stage 4 1/2" in Kohlberg (1973) was Case 65. Table 13 indicates that Case 65 was scored as Stage 3 (Standard Issue Scoring) at the end of high school. In college his scores were Stage 4 on Form A, Stage 2 on Form B, and Stage 2(3) on Form C. Like Case 2, spontaneous college interview material also indicated moral cynicism and relativism. Unlike Case 2, however, Case 65's scattered college scores, including Stage 2 judgments, did not represent a transition from conventional to principled or post-conventional reasoning. Before his college moral "crisis", he had reasoned primarily at Stage 3. After this crisis, his reasoning was successively Stage 4(3) and Stage 4.

Our current interpretation of the phenomenon of "college-age relativism" is that it does not necessarily represent a transitional period between Stages 4 and 5. Instead, moral relativism appears to be a metaethical position that can be taken at a number of different developmental stages. Whereas the level of sophistication of the relativistic metaethin may differ developmentally, the effect of this stance on normative judgments appears to be very much the same whether the subject is moving from Stage 4 to Stage 5 or is shifting within the conventional level. The increased stage scatter in these interviews can be interpreted as reflecting a breakdown of faith in the currently held moral frame of reference with no substitution of a more adequate system. Subjects in this position sometimes refuse to make moral judgments. When they do make moral judgments they seem to rely on whatever has some lingering validity for them even if these intuitions from Stages 2, 3, and 4 fail to represent a coherent whole.

The Stage Model

Let us turn now to the more usual pattern of moral judgment development. Our data seem to provide strong support for the strict Piagetian stage model of development. With regard to the criterion of "structured wholeness" or internal consistency, we found that the great majority of our interviews were scored at only one moral judgment stage or at most two adjacent stages. Only 9% showed evidence of a third stage. On the average, across nine independently scored dilemmas, two-thirds of all scores were assigned at the subject's modal stage and almost all the reamining scores (32%) were assigned to a single stage adjacent to the modal stage. The structured wholeness assumption was also supported by the high degree of alternate form and test-retest reliability, the high Cronbach's alpha, and the results of factor analyses of issue and dilemma scores.

The factor analyses indicate that there was a single factor of general moral level across the domain of dilemmas and issues. The absence of issue or dilemma factors along with the absence of scatter across more than two adjacent stages indicates that we have succeeded in defining a coherent moral domain united by a single underlying organizational structure. This position contrasts with that taken by Damon (1977) who argues that, at least in young children, the moral domain is broken down into a number of issues or dimensions each with its own unique structure. The two moral issues on which Damon has focused are authority and distributive justice. Although Damon's research does not use Kohlberg's hypothetical dilemmas, it is clear that Kohlberg dilemmas do involve both authority and distributive justice. If Damon's interpretation of the moral issues as independently organized were true at the age level addressed by our study, the factor analysis should have indicated multiple factors rather than a single general factor of moral stage that cuts across the content of the dilemmas.

This high degree of structural coherence within each interview leads to a relatively saltatory picture of development if we compare interviews across time for any single individual. developmental change in any given three year period seems to involve predominantly an increase in the next higher stage of reasoning along with a simultaneous decrease in the lower stage. This means that the subject is most often in transition between two adjacent stages, has given up using all earlier stages and uses no reasoning more than one stage above his modal stage. This contrasts with a model in which the distribution of a subject's reasoning extends across all five stages and in which development involves the gradual increase in usage of the higher stages along with decrease (but not decrease to zero) in use of lower stages. Kohlberg's early results based on the Sentence Rating method may have appeared to support this latter model but we now interpret those results as artifactual, i.e., due to inappropriate definition of the scoring unit and a confusion between moral judgment content and structure.

Our results also support the Piagetian stage model's assumption of invariant developmental sequence. With a few exceptions which can be attributed to scoring error, each of our subjects proceeded through the stages in the prescribed order, neither skipping stages nor regressing to an earlier level once a later stage had been attained.

This stage model of development has recently been criticized--in the cognitive domain by Flavell (1980) and, most directly, in the moral judgment domain by Rest (1979). Rest agrees with Kohlberg that qualitatively different forms of moral judgment can be identified and that development involves the increasing use of more advanced or sophisticated types and the decreasing use of less sophisticated types. He disagrees, however, with Kohlberg's claim that development proceeds through a stepwise sequence of internally consistent stages. He holds instead that individuals simultaneously use reasoning of many types and that an adequate description of an individual's moral judgment must include a quantitative account of the proportion of each type rather than a global stage designation for the individual. He interprets our finding of internal consistency as a methodological artifact and points to his own results with the Defining Issues Test as consistent with a more complex model of development.

One interpretation of the discrepancy between Rest's results and our own derives from the fact that the DIT is measuring comprehension and preference of moral judgments made by others while the Standard Issue System is measuring an individual's spontaneous production of moral judgments in response to very open-ended questions. What we are suggesting is that the development of moral judgment as a whole (including comprehension and preference as well as spontaneous production) may be too broad a scope for what Rest calls the "simple stage model". It may be the case that development of moral judgment comprehension and preference does not follow a stage-like pattern but that spontaneous production does when assessment conditions pull for the subject's competence.

This would mean that the developing individual has one basic framework for resolving moral dilemmas and justifying those resolutions. This framework has some coherence for the individual and it undergoes developmental transformation, hierarchically integrates the insights from lower levels, and so on. On the other hand, other aspects of moral development such as the comprehension of judgments made by others may not be stage-like in this sense. In fact, given Kohlberg's assumptions, there is little reason to expect that subjects will comprehend statements at only one stage (or at two adjacent stages). Rather, they can be expected to comprehend all stages below their own as well as possibly a stage or two above their own. Thus we would not characterize the development of moral judgment comprehension as following a stage-like sequence. Preference for moral judgments made by others is somewhat more problematic, however. According to a simple stage model, how would one explain the endorsement of judgments that represent stages below one's own level of spontaneous production? One possible explanation derives from the Piagetian assumption of hierarchical integration. That is, while the stage as a whole may undergo transformation as the individual develops, many of the insights achieved at the lower stages remain valid from the higher stage perspective though they are now imbedded in a more complex and sophisticated position. For example, the Stage 3 judgment that "you ought to keep a promise because the person you promised trusts you" would be endorsed or agreed to by individuals who reason predominantly at Stages 4 and 5 as well as Stage 3. The difference is that at the higher stages one's conception of trust and reasons for its importance have developed beyond what is available at Stage 3.

Another consideration with regard to Rest's preference data is methodological. When subjects endorse an item on the DIT, it is not clear how they understand the item. The Standard Issue Moral Judgment Interview involves probing for elaboration and clarification and pooling all responses to the dilemma which refer to a single idea (as defined by norm and element). This system will naturally lead to greater internal consistency than assigning a score for each unelaborated statement the subject judges to have some validity. We do not interpret our finding as an artifact. It is true, as Rest point out, that according to Standard Issue Scoring rules, an interview cannot receive more than one stage (or stage transition) score for a single norm/element intersection. However, scores at all five stages may be assigned to a single issue, dilemma, or interview if interview/manual matches at those stages occur. In fact, such variance does not occur in our data even though the nine dilemmas were scored independently and by three different raters. We agree with Rest that moral judgment comprehension and preference are important aspects of moral development that are separable from spontaneous moral judgment production and that a full developmental account of an individual's moral judgment requires scores on all those aspects (that is, the information from the various sources is not redundant). However, we feel that a careful look at our methods and data should convince the reader of the validity of the Piagetian stage model for describing the development of spontaneous moral judgment at least within the limits of our instrument and sample.

In general, then, we interpret the results of this study as consistent with cognitive-developmental theory, in particular as consistent with a stage model of development. The results in relation to moral judgment compare favorably with longitudinal studies of the stage model in purely cognitive development (Kohn, 1969). Our subjects did seem to use a coherent structural orientation in thinking about a variety of moral dilemmas. Their thinking developed in a regular way up the stage sequence, neither skipping stages nor reverting to use of a prior stage. Our results also serve to validate the moral judgment stages as operationally defined in Standard Issue Scoring and to indicate that the Standard Issue system is a reliable and valid measure of moral judgment.

Appendix A

Form A

Moral Judgment Interview

Dilemma III: In Europe, a woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost him to make. He paid $200 for the radium and charged $2,000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about $1,000, which is half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So Heinz gets desperate and considers breaking into the man's store to steal the drug for his wife.

1. Should Heinz steal the drug?

1a. Why or why not?

2. If Heinz doesn't love his wife, should he steal the drug for her?

2a. Why or why not?

3. Suppose the person dying is not his wife but a stranger. Should Heinz steal the drug for the stranger?

3a. Why or why not?

4. (If you favor stealing the drug for a stranger:) Suppose it's a pet animal he loves. Should Heinz steal to save the pet animal?

4a. Why or why not?

5. Is it important for people to do everything they can to save another's life?

5a. Why or why not?

6. It is against the law for Heinz to steal. Does that make it morally wrong?

6a. Why or why not?

7. Should people try to do everything they can to obey the law?

7a. Why or why not?

7b. How does this apply to what Heinz should do?

Dilemma III': Heinz did break into the store. He stole the drug and gave it to his wife. In the newspapers the next day, there was an account of the robbery. Mr. Brown, a police officer who knew Heinz, read the account. He remembered seeing Heinz running away from the store and realized that it was Heinz who stole the drug. Mr. Brown wonders whether he should report that Heinz was the robber.

1. Should Officer Brown report Heinz for stealing?

1a. Why or why not?

2. Officer Brown finds and arrests Heinz. Heinz is brought to court, and a jury is selected. The jury's job is to find whether a person is innocent or guilty of committing a crime. The jury finds Heinz guilty. It is up to the judge to determine the sentence. Should the judge give Heinz some sentence, or should he suspend the sentence and let Heinz go free?

2a. Why?

3. Thinking in terms of society, should people who break the law be punished?

3a. Why or why not?

3b. How does this apply to how the Judge should decide?

4. Heinz was doing what his conscience told him when he stole the drug. Should a law-breaker be punished if he is acting out of conscience?

4a. Why or why not?

(Questions 5-10 are designed to elicit the subject's theory of ethics and should be considered optional.)

5. What does the word conscience mean to you, anyhow? If you were Heinz, how would your conscience enter into the decision?

6. Heinz has to make a moral decision. Should a moral decision be based on one's feelings or on one's thinking and reasoning about right and wrong?

7. Is Heinz's problem a moral problem? Why or why not?

7a. In general, what makes something a moral problem or what does the word moral mean to you?

8. If Heinz is going to decide what to do by thinking about what's really right, there must be some answer, some right solution. Is there really some correct solution to moral problems like Heinz's? When people disagree is everybody's opinion equally right? Why?

9. How do you know when you've come up with a good moral decision? Is there a way of thinking or method by which one can reach a good or adequate decision?

10. Most people believe that thinking and reasoning in science can lead to a correct answer. Is the same thing true in moral decisions or are they different?

Dilemma I: Joe is a 14 year old boy who wanted to go to camp very much. His father promised him he could go if he saved up the money for it himself. So Joe worked hard at his paper route and saved up the $40.00 it cost to go to camp and a little more besides. But just before camp was going to start, his father changed his mind. Some of his friends decided to go on a special fishing trip, and Joe's father was short of the money it would cost. So he told Joe to give him the money he had saved from the paper route. Joe didn't want to give up going to camp, so he thinks of refusing to give his father the money.

1. Should Joe refuse to give his father the money?

1a. Why or why not?

2. Is the fact that Joe earned the money himself the most important thing in the situation?

2a. Why or why not?

3. The father promised Joe he could go to camp if he earned the money. Is the fact that the father promised the most important thing in the situation?

3a. Why or why not?

4. Is it important to keep a promise?

4a. Why or why not?

5. Is it important to keep a promise to someone you don't know well and probably won't see again?

5a. Why or why not?

6. What do you think is the most important thing a son should be concerned about in his relationship to his father?

6a. Why is that the most important thing?

7. What do you think is the most important thing a father should be concerned about in his relationship to his son?

7a. Why is that the most important thing?

Dilemma IV: There was a woman who had very bad cancer, and there was no treatment known to medicine that would save her. Her doctor, Dr. Jefferson, knew that she had only about six months to live. She was in terrible pain, but she was so weak that a good dose of a pain-killer like ether or morphine would make her die sooner. She was delirious and almost crazy with pain, and in her calm periods she would ask Dr. Jefferson to give her enough ether to kill her. She said she couldn't stand the pain and she was going to die in a few months anyway. Although he knows that mercy-killing is against the law, the doctor thinks about granting her request.

1. Should Dr. Jefferson give her the drug that would make her die?

1a. Why or why not?

2. Should the woman have the right to make the final decision?

2a. Why or why not?

3. The woman is married. Should her husband have anything to do with the decision?

3a. Why or why not?

4. Is there any way a person has a duty or obligation to live when he or she does not want to, when the person wants to commit suicide?

4a. Why or why not?

5. It is against the law for the doctor to give the woman the drug. Does that make it morally wrong?

5a. Why or why not?

6. Should people try to do everything they can to obey the law?

6a. Why or why not?

6b. How does this apply to what Dr. Jefferson should do?

Dilemma IV': Dr. Jefferson did perform the mercy-killing by giving the woman the drug. Passing by at the time was another doctor, Dr. Rogers, who knew the situation Dr. Jefferson was in. Dr.Rogers thought of trying to stop Dr. Jefferson, but the drug was already administered. Dr. Rogers wonders whether he should report Dr. Jefferson.

1. Should Dr. Rogers report Dr. Jefferson?

1a. Why or why not?

2. The doctor does report Dr. Jefferson. Dr. Jefferson is brought to court and a jury is selected. The jury's job is to find whether a person is innocent or guilty of committing a crime. The jury finds Dr. Jefferson guilty. It is up to the judge to determine the sentence. Should the judge give Dr. Jefferson some sentence, or should be suspend the sentence and let Dr. Jefferson go free?

2a. Why?

3. Thinking in terms of society, should people who break the law be punished?

3a. Why or why not?

3b. How does this apply to how the Judge should decide?

4. The jury finds Dr. Jefferson legally guilty of murder. Would it be wrong or right for the judge to give him the death sentence (a legally possible punishment)? Why?

5. Is it ever right to give the death sentence? Why or why not? What are the conditions when the death sentence should be given (if ever) in your opinion?

Why are these condidtions important?

6. Dr. Jefferson was doing what his conscience told him when he gave the woman the drug. Should a lawbreaker be punished if he is acting out of conscience?

6a. Why or why not?

(Questions 7-12 are designed to elicit the subjects' theory of ethics and should be considered optional)

7. What does the word conscience mean to you, anyhow? If you were Dr. Jefferson, how would your conscience enter into the decision?

8. Dr. Jefferson has to make a moral decision. Should a moral decision be based on one's feelings or on one's thinking and reasoning about right and wrong?

9. Is Dr. Jefferson's problem a moral problem? Why or why not?

9a. In general, what makes something a moral problem or what does the word morality mean to you?

10. If Dr. Jefferson is going to decide what to do by thinking about what's really right, there must be some answer, some right solution. Is there really some correct solution to moral problems like Dr. Jefferson's, or when people disagree is everybody's opinion equally right? Why?

11. How do you know when you've come up with a good moral decision? Is there a way of thinking or method by which one can reach a good or adequate decision?

12. Most people believe that thinking and reasoning in science can lead to a correct answer. Is the same thing true in moral decisions or are they different?

Dilemma II: Judy was a 12 year old girl. Her mother promised her that she could go to a special rock concert coming to their town if she saved up from babysitting and lunch money for a long time so whe would have enough money to buy a ticket to the concert. She managed to save up the $15.00 the ticket cost plus another $3.00. But then her mother changed her mind and told Judt that she had to spend the money on new clothes for school. Judy was disappointed and decided to go to the concert anyway. She bought a ticket and told her mother that she had only been able to save $3.00. That Saturday she went to the performance and told her mother that she was spending the day with a friend. A week passed without her mother finding out. Judy then told her older sister, Louise, that she had gone to the performance and had lied to her mother about it. Louise wonders whether to tell their mother what Judy did.

1. Should Louise, the older sister, tell their mother tnat Judy had lied about the money or should she keep quiet?

1a. Why?

2. In wondering whether to tell, Louise thinks of the fact that Judy is her sister. Should that make a difference in Louise's decision?

3. Is the fact that Judy earned the money herself the most important thing in this situation?

3a. Why or why not?

4. The mother promised Judy she could go to the concert if she earned the money. Is the fact that the mother promised the most important thing in the situation?

4a. Why or why not?

5. Why in general should a promise be kept?

6. Is it important to keep a promise to someone you don't know well and probably won't see again?

6a. Why or why not?

7. What do you think is the most important thing a daughter should be concerned about in her relationship to her mother?

7a. Why is that the most important thing?

8. What do you think is the most important thing a mother should be concerned about in her relationship to her daughter?

8a. Why is that the most important thing?

Dilemma V: In Korea, a company of Marines was way outnumbered and was retreating before the enemy. The company had crossed a bridge over a river, but the enemy were mostly still on the other side. If someone went back to the bridge and blew it up, with the head start the rest of the men in the company would have, they could probably then escape. But the man who stayed back to blow up the bridge would probably not be able to escape alive; there would be about a 4 to 1 change he would be killed. The captain himself is the man who knows best how to lead the retreat. He asks for volunteers, but no one will volunteer. If he goes himself, the men will probably not get back safely and he is the only one who knows how to lead the retreat.

1. Should the captain order a man to go on this very dangerous mission or should he go himself? Why?

2. What is the best justification for saying it is right to send someone besides himself?

2a. Why or how do you say it is right to save more lives in this case, when it means ordering someone to his death?

3. What is the best, or most important reason for saying it is wrong to send someone else, when ordering someone else will save more lives?

4. Does the captain have the right or the authority to order a man if he thinks it best to? Why?

5. Would a man have the right to refuse such an order? Why?

6. The captain has a family, the men do not. Should that enter into his decision? How?

6a. If he is going to pick someone to go, how should he pick someone? Why?

7. There is some conflict between fairness and survival here. Which is more important or how can he deal with both here? What does fairness mean anyhow and why is it important?

Dilemma VIII: In a country in Europe, a poor man named Valjean could find no work, nor could his sister and brother. Without money, he stole food and medicine that they needed. He was captured and sentenced to prison for six years. After a couple of years, he escaped from the prison and went to live in another part of the country under a new name. He saved money and slowly built up a big factory. He gave his workers the highest wages and used most of his profits to build a hospital for people who couldn't afford good medical care. Twenty years had passed when a tailor recognized the factory owner as being Valjean, the escaped convict whom the police had been looking for back in his home town.

1. Should the tailor report Valjean to the police? Would it be right or wrong to keep it quiet? Why?
2. What would be the best or most important reason for reporting this?
3. What would be the most or most important reason for keeping quiet?
4. If Valjean were reported and brought before the judge, should the judge send him back to jail? Why?
5. Would it be unfair or unjust to send him back to jail or would it still be just? Why?
6. Suppose that Valjean had escaped from jail and lived an ordinary life, instead of building a hospital to help other sick people. Should the tailor report him in that case? Should he be sent to jail? Why?
7. Does a citizen have a duty or obligation to report an escaped convict?
8. According to the law, a citizen is required to report an escaped convict. Is it morally right to fail to report him in that case? Why?
9. Suppose Valjean was a close friend of the tailor. Does that make a difference in what he should do? Why?

VII. Two young men, brothers, had gotten into serious trouble. They were secretly leaving town in a hurry and needed money. Alexander, the older one, broke into a store and stole $500. Joe, the younger one, went to a retired old man who was known to help people in town. Joe told the man that he was very sick and he needed $500 to pay for the operation. Really he wasn't sick at all, and he had no intention of paying the man back. Although the man didn't know Joe very well, he loaned him the money. So

Joe and Alex skipped town, each with $500.

1. If you had to say who did worse, would you say Karl did worse to break in the store and steal the $500 or Bob did worse to borrow the $500 with no intention of paying it back? Why?

2. Which would you rather do, which would make you feel worse stealing like Karl or cheating like Bob?

3. When people say their conscience bothers them what do they mean?

3a. What does someone do when his conscience bothers him?

3b. Do some people have more conscience than others?

3c. Where do you fit in there?

3d. Do you think your conscience should be the basis of making a decision, or should you think of other things?

4. What was the worst thing about what Karl did in stealing? Why?

5. What was the worst thing about what Bob did in cheating the man? Why?

6. Does that enter into deciding which was worse? Why?

7. Why shouldn't someone steal from a store or steal anyhow?

8. Bob deceived the old man and agreed to do something he didn't. Why shouldn't someone do that anyhow?

9. Does the law enter into your decision in this case?

10. Both men are caught and brought before the judge. Should one be given more punishment than the other? Why?

Appendix B

Illustrations of Moral Norms and Elements

The following examples illustrate judgments which exhibit similar content (issue, norm, and element) but different stage structures.

STAGE: 3
NORM: Affiliation
ELEMENT: Having a duty

(1) (Heinz should steal the drug or it is Heinz's duty to steal the drug) because he should care about her, or love her; OR because he is supposed to stick close to her; etc.

(2) (Heinz should steal the drug even if he doesn't love his wife) because he must have loved her at one time or should still care about her; OR because he is still her husband; etc.

STAGE: 4
NORM: Affiliation
ELEMENT: Having a duty

(Heinz should steal the drug) because he is obligated by his marital responsibility, wedding vows, convenant of marriage; etc.

STAGE: 4/5
NORM: Affiliation
ELEMENT: Having a duty

(Heinz should steal the drug) because he freely accepted an explicit responsibility or commitment in marriage and is now obligated to honor it.

Each moral judgment can be seen as an argument for one of the two conflicting issues in the dilemma. In justifying his/her argument for that issue choice, the subject appeals to a moral value which we have called a norm and construes the importance or meaning of that norm in terms of a third kind of value, an element. For example, the following statement argues that Heinz should steal the drug (life issue) because of the value of love or affiliation (the norm) and interprets the value of love in terms of the element of role duties.

NORM: Affiliation
ELEMENT: Duty

(1) (Heinz should steal the drug or it is Heinz's duty to steal the drug) because he should care about her, or love her; OR because he is supposed to stick close to her; etc.

(2) (Heinz should steal the drug even if he doesn't love his wife) because he must have loved her at one time or should still care about her; OR because he is still her husband, etc.

In contrast, the following judgment argues that Heinz should steal the drug (life issue) because of the same value or norm - love or affiliation - but interprets the importance of that value in terms of the element of reciprocity or positive deservingness.

NORM: Affiliation
ELEMENT: Reciprocity

(1) (Heinz should steal the drug even if he doesn't love his wife) out of gratitude or appreciation; OR because she has shared her life with him, and the least he can do is to save her; etc.

Both judgments represent the same stage of moral judgment (Stage 3) but they are conceptually distinguishable from one another and can be considered two different "ideas".

Appendix C

Illustration of the Explicated Criterion Judgment

CRITERION JUDGMENT (Form B)

DILEMMA: IV

ISSUE: Law

NORM: Law

ELEMENT: Seeking reward (avoiding punishment)

STAGE: 2

CRITERION JUDGMENT

(The doctor should not give the women the drug) because he would risk losing his job or going to jail.)

(Note: Do not match score this point if it is a response to the general question "Why is it important to obey the law?" unless the response refers to the doctor in this mercy killing situation.)

STAGE STRUCTURE

Not killing the woman is justified because it involves a risk (rather than certainty) of punishment. Punishment is seen as something to be instrumentally avoided. The risk of punishment overrides the recognition of the pragmatic reasonableness from the women's point of view of giving her the drug.

CRITICAL INDICATORS

One of the following must be used as the central justification for not killing the woman: (a) punishment as possible or probable, a risk to be weighed in the decision; or (b) other possible disadvantageous consequences to the doctor (he might lose his job, etc.)

DISTINCTIONS - OTHER STAGES

This Stage 2 treatment of sanctions as a factor to be included among one's pragmatic or instrumental considerations should be distinguished from the Stage 1 view of sanction as automatically defining the wrongness of an act and from the Stage 1/2 focus on the likelihood of punishment as a reason not to mercy kill. In Stage 1/2 judgments, it is ambiguous as to whether punishment automatically defines the act as wrong or whether it is seen as a pragmatic consideration. At Stage 2, the pragmatic quality of the concern is clear.

Do not score as matches the statements that it is unfair of the woman to expect the doctor to risk punishment, that the doctor is not obligated to kill her if he will be risking punishment or loss of license, that he has the right to consider consequences to himself, or that he should not do it if it will mean loss of the right to practice medicine, destruction of his career, etc.

MATCH EXAMPLES

1. SHOULD THE DOCTOR GIVE HER THE DRUG THAT WOULD MAKE HER DIE? WHY?

   No, the doctor could be charged with killing her. He should give something to calm her. (WHY?) He would lose his career and go to prison. He should protect himself first and not kill her.

The following euthanasia dilemma responses from two different interviews both seem to approximate the criterion judgment presented here. In each case, one must decide whether or not the response can be said to embody the concept specified in the criterion judgment. Without the critical indicators this decision is very difficult to make, and the inter-rater reliability for such decisions is not likely to be high. In most cases, the critical indicators resolve such questions fairly unambiguously.

Example 1

SHOULD THE DOCTOR GIVE HER THE DRUG THAT WOULD MAKE HER DIE? WHY?

No. He would be blamed for killing her. She could take her own overdose. If he did, he could lose his license and be out of a job.

Example 2

SHOULD THE DOCTOR DO WHAT SHE ASKS AND GIVE HER THE DRUG?

No, I don't think so. I think it's asking too much of a doctor for one thing, that she should ask this even though she is in great pain. A doctor isn't supposed to do this. WHY? I believe it's in their code that you shouldn't give a drug to any person to help them die sooner or to put them to death right away. If he were found out to have given her this drug, he'd probably be kicked out of his profession and he might not be able to get into something else. (Case 41-B)

This (Form B) dilemma involves the question of whether a mercy killing should be performed for a woman who is terminally ill, in great pain and asking to be killed. At first glance these two responses both seem to warrant a score of Stage 2 on the basis of this criterion judgment. However, only the first qualifies as a match based on the critical indicators.

The second example refers to the likelihood that the doctor will lose his job as required by critical indicator (b), but the risk of undesirable consequences is not used as the central argument against mercy-killing as the critical indicators specify. The requirement that the risk of losing his job be used as a central justification reflects the fact that at higher stages a pragmatic concern such as this maintains some validity (the stages are "hierarchically integrated") but is no longer seen as a sufficient justification from the moral point of view. In fact, longitudinal case 41 (Time B) was scored globally at Stage 3.

Scoring for more detailed, concrete, and narrowly defined concepts rather than for the stage structures in their more general form has also made it possible to gather evidence for sequentiality and consistency in all aspects of the more detailed stage descriptions. That is, the smaller scoring unit has made it possible to describe the developmental distinctions among parallel "moral ideas" across the stages. For example, in the punishment follow-up to the Heinz dilemma, the idea of deterrance as an argument for punishment is used at seven points in the sequence, beginning at transitional level 1/2 and proceeding through Stages 2, 2/3, 3, 3/4, 4, and 4/5. The criterion judgments specify exactly what the conception of deterrence is at each point, and the "Distinctions" paragraphs explain the differences among the seven interretations.

Appendix D

Standard Issue Scoring Rules

As discussed in Chapter I, Standard Issue Scoring involves classifying responses to a dilemmas into the two standard issue categories and then assigning a stage score for each match between a manual criterion judgment and a moral judgment in the interview.

The only exception to this occurs when a higher stage match constitutes the elaboration of an idea represented in a lower stage match. In this case, only the higher stage score is assigned a subject might, for example, argue that if Heinz steals the drug the judge should punish him because "when someone has broken the law a judge is expected to or supposed to give a sentence." Such a statement matches a Stage 3 punishment issue criterion judgment. The subject might then go on to elaborate his/her conception of the judge's role in Stage 4 terms (the judge should punish because he has accepted a responsibility to uphold the laws of society). In such a case only a Stage 4 score would be assigned. If, however, the subject made the Stage 3 statement referred to above along with a Stage 4 statement that did not constitute an elaboration of the Stage 3 idea (e.g. the judge should punish Heinz in order to promote respect for property rights), then two scores would be assigned - one at Stage 3 and one at Stage 4.

Sometimes an idiosyncratic or poorly probed interview will not yield even a single match for the issue being scored. In such cases, if moral judgment material is found corresponding to the issue in question, the material is assigned a "guess score" which represents the rater's intuitive evaluation of the stage of that material.

In calculating an overall stage score for a three-dilemma interview, one begins by assigning a score for each of the six issues based on the criterion judgment match scores (or guess scores) assigned for material on each issue. As many as three different stages may be represented in the issue total. In our data we have never encountered an interview with matches at more than three stages for a single issue. Even the case of three stages on a single issue almost never occurs. The six issue scores are then combined (with guess-issue scores given half the weight of issue scores based on matches) to yield a global score and a "moral maturity" score for the interview. The global score consists of a major stage (the subject's modal level) and one or more minor stages (any stage that has received at least 25% of the total points assigned the interview). A global score of Stage 3, for example, indicates that most of the scores assigned the interview were Stage 3 and that no stage other than Stage 3 reached the 25% cutoff point. A global score of Stage 3(2) indicates that the subject's modal stage was Stage 3 and that s/he also received at least 25% of his/her (weighted) scores at Stage 2. The moral maturity score represents a weighted average of the issue scores and ranges from 100 (Stage 1) to 500 (Stage 5).

Computation Instructions for Global Scores and MMS

Computing Issue Scores

1. A three-class scale is sued for issue scores. That is, an individual issue score can be either a pure stage score, a major-minor score, or a transitional score, e.g., 2, 2(3), 2-3, 3(2) or 3.
2. Where only one stage is represented in the criterion judgment scores for an issue or where only one stage exceeds 25% of the total criterion judgment scores, that stage score is the issue score. That is, a pure stage score is given for the issue.
3. If two or more stages are represented each by 25% or more of the issue's criterion judgment scores, all are registered in the issue total with the one used most frequently being the major stage, the ones used less frequently the minor stages. (They should appear in order of frequency.) If the frequency of the two (or three) is exactly equal, a transitional score with no major-minor designation is assigned. Scores for chosen issues should be circled unless they are guess scores. Guess scores are not circled.

Global and MMS

In computing global and moral maturity scores, use the following weights:

3 Chosen issue (based on match scores)

2 Non-chosen issue (based on match scores)

1 Guess score (whether on chosen or non chosen issue).

When an issue score is a pure stage, assign 3 points to the stage if it's a non-chosen issue (match), and one if it's a guess.

When the score for a chosen issue involves a major and a minor stage (e.g., 3(2), assign two of the three points to the major stage, one point to the minor stage. When the score for a chosen issue is a transitional stage (e.g., 2/3), assign one and a half points to each stage.

When a non-chosen issue involves a major and a minor stage, assign one point for the major and one point for the minor stage. That is, treat major/minor scores on non-chosen issues as if they were transitional type scores.

When a guess score involves two stages (e.g., G2(3) or G2/3), assign one half point for each stage.

Begin by listing the stages which appear in the issue totals. For example:

3 Life
2(1) Morality and Conscience
G3 Contract
2(3) Law
G2/3 Punishment
2/3 Authority

This interview involves the use of three stages -- 1, 2, and 3.

Stage 1	1
Stage 2	5
Stage 3	7
Total	13

Stage 1 is assigned only one point because it occurs only once, as the minor stage of a chosen issue score (2(1)). The major stage of this issue, Stage 2, receives two points. Stage 2 also receives one point for the law issue (half of the two points given to a non-chosen match based score), 1/2 point for the punishment issue (half of the one point given for a guess issue score), and 1 1/2 points for the authority issue (half of the three points for this chosen issue). The total for Stage 2 is five points. Stage 3 receives three points for life (chosen issue), one point for law, one half point for pun, one point for contract, and one and one-half points for authority for a total of seven points. The total number of points assigned for the interview is 13.

Global stage is assigned as follows:

The stage with the largest proportion of points is the major stage. If a second stage has received at least 25% of the total assigned points, it is entered as a minor stage. If two stages exceed 25% of the total, two minor stages are listed in order of proportion of points assigned the stages. If two stages (or three) are assigned an exactly equal number of points, the highest stage is the major stage. For example, an interview with 50% at Stage 2 and 50% at Stage 3 would be assigned a global score of 3(2).

Clearly in our example, Stage 3 is the major stage for the interivew since it received the most points. Stage 2 is entered as a minor stage since it exceeds 25% of the total (5/13 = 38%). Stage 1 is not entered since it received less than 25% of the total (1/13).

Moral Maturity Score

In order to calculate MMS, multiply each stage by the number of points at that stage:

$$\begin{array}{ccccc} 1 & \times & 1 & = & 1 \\ 2 & \times & 5 & = & 10 \\ 3 & \times & \underline{7} & = & \underline{21} \\ & & 13 & & 32 \end{array}$$

Divide the sum of the products (32) by the total number of points assigned (13) and then mulitply by 100.

$$\begin{array}{r} 2.46 \\ 13\overline{)32} \end{array} \times 100 = 246.$$

References

Belenky, M. Conflict and Development: A Longitudinal Study of the Impact of Abortion Decisions on Moral Judgments of Adolescent and Adult Women. Unpublished doctoral dissertation, Harvard University, 1978.

Blasi, A. Bridging moral cognition and moral action: A critical review of the literature. Psychological Bulletin, 1980, 88,1,1-45.

Blatt, M. and Kohlberg, L. The effects of classroom moral discussion upon children's level of moral judgment. Journal of Moral Education, 1975, 4, 129-163.

Bloom, B. Stability and change in human characteristics. New York: Wiley, 1964.

Colby, A. Evolution of a moral-developmental theory. In W. Damon (Ed.), New Directions for Child Development: Moral Development, 2, San Francisco: Jossey-Bass, 1978.

Colby, A., Kohlberg, L., Gibbs, J., Candee, D., Hewer, A., Power, C. and Speicher-Dubin, B. Measurement of Moral Judgment: Standard Issue Scoring Manual. New York: Cambridge University Press, in press.

Colby, A., Kohlberg, L., Fenton, E., Speicher-Dubin, B., and Lieberman, M. Secondary school moral discussion programmes led by social studies teachers. Journal of Moral Education, 6, 2, 1977.

Damon, W. The Social World of the Child. San Francisco; Jossey-Bass, 1977.

Damon, W. Social Development from Childhood through Adolescence. New York: Norton, in press.

Edwards, C. P. The comparative study of the development of moral judgment and reasoning. In Munroe, Munroe, and B. Whiting (Eds.), Handbook of Cross-Cultural Human Development. N.Y.: Garland Press, 1981.

Erickson, V. L. The case study method in the evaluation of developmental programs. In L. Kuhmerker, M. Mentkowski, and L. V. Erickson (eds.), Evaluating Moral Development. New York: Character Research Press, 1980.

Flavell, J. J. Stage-related properties of cognitive development. Cognitive Psychology, 1971, 2, 521-453.

Gilligan, C. In a different voice. Harvard Educational Review, 1977, 47(3), 365-378.

Gilligan, C. and Belenky, M. A naturalistic study of abortion decisions. In R. Selman and R. Yando (Eds.), Clinical-Developmental Psychology. San Francisco: Jossey-Bass, 1980, No.7.

Gilligan, C. and Murphy, J.M. Development from adolescence to adulthood: the philosopher and the dilemma of the fact. In D. Kuhn (Ed.), Intellectual Development beyond Childhood. San Francisco: Jossey-Bass, 1979, No.5.

Gibbs, J. Kohlberg's moral stage theory: A Piagetian revision. Human Development, 1979, 22, 89-112.

Holstein, C. Irreversible, stepwise sequence in the development of moral judgment: a longitudinal study of males and females. Child Development, 1976, 47, 51-61.

Kohlberg, L. The development of modes of moral thinking and choice in the years ten to sixteen. Unpublished doctoral dissertation, Univ. of Chicago, 1958.

Kohlberg, L. Stage and sequence: The cognitive-developmental approach to socialization. In D.A. Goslin (Ed.) Handbook of Socialization Theory and Research. Chicago: Rand McNally, 1969.

Kohlberg, L. The moral atmosphere of the school. In N. Overley (Ed.) The Unstudied Curriculum, Monograph of the Association for Supervision and Curriculum Development. Washington, D.C. 1970.

Kohlberg, L. Continuities in childhood and adult moral development revisited. In P. Baltes and K.W. Schaie (Eds.), *Life Span Developmental Psychology*. New York: Academic Press, 1973.

Kohlberg, L. Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral Development and Behavior*. New York: Holt, Rinehart and Winston, 1976.

Kohlberg, L. *The meaning and measurement of moral development*. Worcester, MA: Clark University Press, 1980.

Kohlberg, L., and Candee, D. The relation between moral judgment and moral action. Paper presented at the International Conference on Morality and Moral Development, Florida International University, Miami, December, 1981.

Kohlberg, L. and Kramer, R. Continuities and discontinuities in childhood and adult moral development. *Human Development*, 1969, *12*, 93-120.

Kohn, N. Performance of negro children of varying social class background on Piagetian tasks. Unplublished doctoral dissertation. Univ. of Chicago, 1969.

Kramer, R. Moral development in young adulthood. Unpublished doctoral dissertation, Univ. of Chicago, 1968.

Kuhn, D. Short-term longitudinal evidence for the sequentiality of Kohlberg's early stages of moral judgment. *Developmental Psychology*, 1976, *12*, 2, 162-166.

Kurtines, W. and Greif, E. The development of moral thought: Review and evaluation of Kohlberg's approach. *Psychological Bulletin*, 1974, 81, 8, 453-470.

Lakatos, I. Falsification and Methodology of Scientific Research Programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1976.

Lockwood, A. L. The effects of values clarification and moral development curriculum on school-age subjects: a critical view of recent research. Unpublished manuscript, University of Wisconsin, 1977.

Loevinger, J. Ego Development, San Francisco. Jossey-Bass, 1976.

Loevinger, J. and Wessler, R. Measuring Ego Development. San Franscisco: Jossey-Bass, 1978.

Mead, G. H. Mind, Self, and Society. Chicago: University of Chicago Press, 1934.

Murphy, J. M. An alternative construction of adult moral development. Final Report for NIMH Grant RO3MH31571, 1980.

Nicolayev, J. and Phillips, D.C. On assessing Kohlberg's stage theory of moral development. In D. Cochrane, C. Hamm, and A. Kazepides (Eds.), The Domain of Moral Education. New York: Paulist Press, 1979.

Nisan, M., and Kohlberg, L. Universality and variation in moral judgment -- a longitudinal and cross-sectional study in Turkey. Child Development, in press.

Piaget, J. The moral judgment of the child. New York: Free Press, 1965. (Originally published in English, London: Kegan Paul, 1932).

Puka, B. A Kohlbergian reply. In D. Cochrane, C. Haam, and A. Kazepides (Eds.) The Domain of Moral Education. New York: Paulist Press, 1979.

Rest, G. Voting preference in the 1976 Presidential Election and the influences of moral reasoning. Unpublished manuscript, University of Michigan, 1977.

Rest, J. The hierarchical nature of moral judgment: A study of patterns of comprehension and preference of moral stages. Journal of Personality, 1973, 41, 86-109.

Rest, J. Morality. In J. Flavell and C. Markman (Eds.), Carmichael's Manual of Child Psychology, 4th edition, in press.

Rest, J. Development in Judging Moral Issues. Minneapolis: Univ. of Minnesota Press, 1979.

Selman, R. and Jaquette, D. Stability and oscillation in interpersonal awareness: A clinical-developmental analysis. In C.B. Keasey (Ed.), Twenty-fifth Nebraska Symposium on Motivation. Lincoln: University of Nebraska Press, 1978.

Snarey, J. The moral development of Kibbutz founders and sabras: A cross-sectional and ten year longitudinal cross-cultural study. Unpublished doctoral dissertation, Harvard University, 1982.

Turiel, E. Social-cognitive development: Domains and categories. Paper prepared for a plenary session of the Ninth Annual Symposium of the Jean Piaget Society. Philadelphia, 1979.

White, C. B., Bushnell, N., and Regnemer, J.L. Moral development in Bahamian school children: A three-year examination of Kohlberg's stages of moral development. Developmental Psychology, 1978, 14, 58-65.

Acknowledgements

We wish to acknowledge our debt to Dan Candee, Ros Hewer, Clark Power, and Betsy Speicher-Dubin for their years of arduous and careful labor in refining the moral stage definitions, creating the Standard Issue Scoring System, and scoring hundreds of moral judgment interviews. Ann Higgins, Kelsey Kauffman, and Laura Rogers also contributed in important ways to the development of the stage definitions and scoring system. We acknowledge also the patient efforts of Dan Candee, Bill Jennings, and Steve Reese in tracking down and interviewing our subjects. We would like to thank Bill Damon, Jane Loevinger, Bill Puka, and Jim Rest for reading earlier drafts of the manuscript and offering insightful comments and suggestions which helped to shape the monograph in its final form. We are indebted to the National Insitute of Child Health and Human Development for generous funding of this research. Finally, we are most grateful to the 58 men who contributed their time and thoughtful responses to this project over a 20 year segment of their lives. Without these men, our research subjects, the project would, indeed, have been impossible.

Footnotes

1. The research reported here was supported by National Institute of Child Health and Human Development Grants # HD02469-01 and 5R01HD04128-09.
2. We will briefly address the issue of circularity later in this paper. For a fuller discussion of that issue and of the theoretical and methodological changes referred to here, see Colby (1978).
3. Each case included four to six interviews with the same subject collected at 4 year intervals.
4. All of the data from Kohlberg's longitudinal study are archived at the Henry A. Murray Research Center of Radcliffe College where they are available to interested researchers for further analysis.
5. Girls were not included in Kohlberg's original sample because adding gender as a fourth variable would have required doubling the sample. Given the laboriousness of the interviewing and scoring procedures, such a large sample was not feasible. In retrospect, however, the omission of girls is regrettable.
6. Note that we treat reasoning that represents less than 10% of the subject's total as error. For example, the 4% of Stage 3 at age 10 reflects less than one full CJ across all dilemmas scored and thus should not be considered to be true use of Stage 3 reasoning.
7. Recall that the test-retest correlations were in the high 90's. This indicates that the unexplained variance in predicting from earlier to later ages cannot be attributed to measurement error.
8. The Stage 5 present in the low SES group is due to one subject (Case 2)

reaching Stage 4/5 at age 24. Since he was age 10 at Time 1 and age 28 at Time 6, he was not interviewed at age 32 or 36. That is, the decrease in mean percentage of Stage 5 does not reflect a decrease in Stage 5 use in any one subject, but rather the absence of the one Stage 4/5 subject at the last two ages.

9. The scoring manual itself is currently in press with Cambridge University Press. Consistent use of the published manual should make possible greater comparability across independently conducted studies.

10. One need not accept Phillips' and Nicolayev's statement that "the assumption that moral development occurs in fixed stages...could not be given up, for then nothing of the program would remain." Although we consider the stage concept to be central, others, such as Rest (1979), do not.

11. Only the middle class subsample is used because a comparable figure for the total sample based on Sentence Rating scores was not available.

12. Kohlberg and Kramer analyzed interviews of some longitudinal subjects as old as age 24 as well as the moral judgment interviews of their fathers.

13. Rest's Defining Issues Test involves the presentation of hypothetical moral dilemmas along with a set of stage keyed ideas, issues, or considerations relevant to resolution of each dilemma. The subject is asked to rate and rank the importance of these considerations in his thinking about the dilemma. Scores are based on the relative importance for the subject of considerations at the six moral judgment stages.

Table 1

Six Stages of Moral Judgment

(Reprinted from Kohlberg, 1976)

Level and Stage	Content of Stage What is Right	 Reasons for Doing Right	 Sociomoral Perspective of Stage
LEVEL 1 - PRECONVENTIONAL Stage 1 - Heteronomous Morality	To avoid breaking rules backed by punishment, obedience for its own sake, and avoiding physical damage to persons and property.	Avoidance of punishment, and the superior power of authorites.	Egocentric point of view. Doesn't consider the interests of others or recognize that they differ from the actor's; doesn't relate two points of view. Actions are considered physically rather than in terms of psychological interes of others. Confusion of authorit perspective with one's own.
Stage 2 - Individualism, Instrumental Purpose, and Exchange	Following rules only when it is to someone's immediate interest; acting to meet one's own interests and needs and letting others do the same. Right is also what's fair, what's an equal exchange, a	To serve one's own needs or interests in a world where you have to recognize that other people have their interests, too.	Concrete individualistic perspect Aware that everybody has his own interests to pursue and these conflict, so that right is relative (in the concrete individualistic sense).

Table 1 (continued)

LEVEL II-CONVENTIONAL Stage 3 - Mutual Interpersonal Expectations, Relationships, and Interpersonal Conformity	Living up to what is expected by people close to you or what people generally expect of people in your role as son, brother, friend, etc. "Being good" is important and means having good motives, showing concern about others. It also means keeping mutual relationships, such as trust, loyalty, respect and gratitude.	The need to be a good person in your own eyes and those of others. Your caring for others. Belief in the Golden Rule. Desire to maintain rules and authority which support stereotypical good behavior.	Perspective of the individual in relationships with other individuals. Aware of shared feelings, agreements, and expectations which take primacy over individual interests. Relates points of view through the concrete Golden Rule, putting yourself in the other guy's shoes. Does not yet consider generalized system perspective.

Table 1 (continued)

Stage 4 - Social System and Conscience	Fulfilling the actual duties to which you have agreed. Laws are to be upheld except in extreme cases where they conflict with other fixed social duties. Right is also contributing to society, the group, or institution.	To keep the insitution going as a whole, to avoid the breakdown in the system "if everyone did it," or the imperative of conscience to meet one's defined obligations.	Differentiates societal point of view from interpersonal agreement or motives. Takes the point of view of the system that define roles and rules. Considers individual relations in terms of plac in the system.

Table 1 (continued)

LEVEL III - POST-CONVENTIONAL, or PRINCIPLED Stage 5 - Social Contract or Utility and Individual Rights	Being aware that people hold a variety of values and opinions, that most values and rules are relative to your group. These relative rules should usually be upheld, however, in the interest of impartially and because they are the social contract. Some nonrelative values and rights like life and liberty, however, must be upheld in any society and regardless of majority opinion.	A sense of obligation to law because of one's social contract to make and abide by laws for the welfare of all and for the protection of all people's rights. A feeling of contractual commitment, freely entered upon, to family, friendship, trust and work obligations. Concern that laws and duties be based on rational calculation of overall utility, "the greatest good for the greatest number."	Prior-to-society perspective. Perspective of a rational indivi ual aware of values and rights prior to social attachments and contracts. Integrates perspecti by formal mechanisms of agreeme contract, objective impartialit and due process. Considers mora and legal points of view; recognizes that they sometimes conflict and finds it difficul to integrate them.

Table 1 (continued)

Stage 6 - Universal Ethical Principles	Following self-chosen ethical principles. Particular laws or social agreements are usually valid because they rest on such principles. When laws violate these principles. one acts in accordance with the principle. Principles are universal principles of justice: the equality of human rights and respect for the dignity of human beings as individual persons.	The belief as a rational person in the validity of universal moral principles, and a sense of personal commitment to them.	Perspective of a moral point of view from which social arrangements derive. Perspective is that of any rational individual recognizing the nature of morality or the fact that persons are ends in themselves and must be treated as such.

Table 2

Categories of Moral Content

The Elements

Upholding Normative Order:

1. Obeying (consulting) persons or deity. Should obey, get consent. (should consult, persuade).
2. Blaming (approving). Should be blamed for, disapproved (should be approved).
3. Retributing (exonerating). Should retribute against (should exonerate).
4. Having a right (having no right).
5. Having a duty (having no duty).

Egoistic Consequences:

6. Good reputation (bad reputation)
7. Seeking reward (avoiding punishment)

Utilitarian Consequences:

8. Good individual consequences (bad individual consequences)
9. Good group consequences (bad group consequences)

Ideal or Harmony Serving Consequences:

10. Upholding character
11. Upholding self-respect
12. Serving social ideal or harmony
13. Serving human dignity and autonomy

Fairness:

14. Balancing perspectives or role-taking

Table 2 (continued)

15. Reciprocity or positive desert
16. Maintaining equity and procedural fairness
17. Maintaining social contract or freely agreeing

The Norms

1. Life a) Preservation
   b) Quality/quantity
2. Property
3. Truth
4. Affiliation
(5. Erotic love and sex)
6. Authority
7. Law
8. Contract
(9. Civil rights)
(10. Religion)
11. Conscience
12. Punishment

Table 3

Design of Study

	Lower SES		Higher SES		
	Integrates	Isolates	Integrates	Isolates	Totals
10 yrs.	6 (6)	4 (6)	6 (6)	5 (6)	21
13 yrs.	4 (6)	3 (6)	5 (6)	5 (6)	17
16 yrs.	6 (6)	2 (6)	3 (6)	4 (6)	14
Totals	16	9	14	14	
	25		28		
	53				

(Figures not in parentheses indicate number of cross-sectional subjects who were followed up longitudinally. Figures in parentheses indicate number of subjects in the original cross-sectional sample. As noted in text, five additional working class subjects, aged 19 at Time 3, were added in 1964 (Time 3).)

Table 4

Religious Composition of Cross-sectional Sample

Working class group	
Catholic	14
Protestant	22
Upper middle-class group	
Catholic	3
Jewish	2
Protestant	31

Table 5

Frequency of Longitudinal Follow-up

Number of interviews completed and scored	N
2	3
3	8
4	25
5	12
6	10

Total N = 58[a]

---

[a] Because interviews from seven of these subjects were used to construct the scoring manual, the blind sample used for data analysis included only 51 subjects.

Table 6

Time 1 Moral Maturity Scores for Drop-outs from Sample

Age at Time 1	10	13	16
Mean MMS at Time 1 of S's			
Dropping out after Time 2	203 (2)	-	-
Dropping out after Time 3	-	-	-
Dropping out after Time 4	176 (4)	240 (2)	260 (5)
Dropping out after Time 5	191 (5)	227 (2)	249 (5)
Mean MMS at Time 1 of drop-outs	190 (11)	234 (4)	255 (10)
Mean MMS at Time 1 of total sample	189 (21)	236 (17)	262 (14)

Table 7

Test-retest Reliability

Correlation $T_1 - T_2$: Form A: .96 (Rater 1); .99 (Rater 2)

Form B: .97 (Rater 2)

Percent agreement within one-third stage:		
	Form A	93%
	Form B	94%
	Form A&B	100%

Percent agreement using pure stage and mixed stage scores

(9 categories: 1, 1/2, 2, 2/3, 3, 3/4, 4, 4/5, 5)

Form A 70% (Rater 1), 77% (Rater 2)

(N=43)

Form B 75% (Rater 2)

(N=31

Forms A&B 80% (Rater 2)

(N=10)

Percent agreement using major/minor stage differentiations

(13 categories: 1, 1(2), 2(1), 2, 2(3), 3(2), 3, 3(4), 4(3), 4, 4(5), 5(4), 5)

Form A 59% (Rater 1), 70% (Rater 2)

Form B 62% (Rater 2)

Form A&B 70% (Rater 2)

Table 8

Inter-rater Reliability

Percent agreement:

		Agreement within 1/3 stage	Complete agreement (9 categories)	Complete agreement (13 categories)
Form A				
Rater pair	1	100	88	53
Rater Pair	2	100	88	63
Rater Pair	3	100	75	63
Rater Pair	4	88	88	63
Rater Pair	5	88	88	63
Form B				
Rater Pair	6	100	78	78
Rater Pair	7	100	88	63
Form C				
Rater Pair	8	91	76	52

Correlation - Raters 1 and 2, Form A test-retest interviews = .98, Form B = .96, Form C = .92

Table 9

Alternate Form Reliability

Longitudinal sample (N = 193)

Correlation Form A Form B = .84

(Rater 1 for Form A, Rater 2 for Form B)

85% agreement within 1/2 stage

(other percent agreement figures not available)

Test-retest sample (Rater 2 for both forms)

Correlation Form A x Form B = .95

Percent agreement (9 categories) = 75%

Percent agreement (13 categories) = 67%

Percent within 1/3 stage = 100%

Table 10

Alternate Form Correlations within Age Groups

Age	10	13-14	16-18	20-22	24-25	28-30	32-33
Form A x Form B							
r =	.37	.39	.73	.65	.78	.78	.68
N =	(19)	(35)	(43)	(29)	(22)	(35)	(22)
Form A x Form C							
r =	.37	.40	.67	.59	.83	.70	.47
N =	(19)	(35)	(42)	(31)	(21)	(26)	(11)
Form B x Form C							
r =	.56	.46	.62	.67	.77	.55	.61
N =	(21)	(35)	(43)	(30)	(21)	(26)	(11)

Table 11

Stage Distributions of Scores within Age Groups for Forms A, B, and C

		Stage								
Age	Form	1	1/2	2	2/3	3	3/4	4	4/5	Mean MMS
10	A	5.3	26.3	42.1	15.8	10.5	0.0	0.0	0.0	204
	B	0.0	52.4	33.3	9.5	4.8	0.0	0.0	0.0	184
	C	4.8	57.1	19.0	19.0	0.0	0.0	0.0	0.0	181
13-14	A	0.0	11.1	8.3	58.3	16.7	2.8	0.0	0.0	249
	B	2.8	5.6	33.3	36.1	11.1	8.3	0.0	0.0	235
	C	0.0	11.1	16.7	55.6	13.9	2.8	0.0	0.0	246
16-18	A	0.0	2.3	4.5	22.7	31.8	36.4	0.0	0.0	299
	B	0.0	4.5	9.1	27.3	16.4	20.5	0.0	0.0	280
	C	2.3	9.1	22.7	36.4	27.3	0.0	0.0	0.0	288
20-22	A	0.0	0.0	0.0	6.5	25.8	54.8	9.7	3.2	335
	B	0.0	0.0	3.3	12.2	40.0	26.7	13.3	3.3	319
	C	0.0	0.0	0.0	18/2	27.3	45.5	9.1	0.0	325
24-26	A	0.0	0.0	0.0	4.3	17.4	43.5	21.7	13.0	365
	B	0.0	0.0	4.2	4.2	25.0	37.5	16.7	8.3	354
	C	0.0	0.0	0.0	4.5	22.7	50.0	18.2	4.5	350
28-30	A	0.0	0.0	0.0	0.0	21.6	48.6	16.2	13.5	362
	B	2.8	0.0	0.0	0.0	13.9	52.8	16.7	13.9	366
	C	0.0	0.0	0.0	0.0	33.3	59.3	7.4	0.0	338

Table 11 (continued)

Stage Distributions of Scores within Age Groups for Forms A, B, and C

		Stage								
Age	Form	1	1/2	2	2/3	3	3/4	4	4/5	Mean MMS
32-33	A	0.0	0.0	0.0	0.0	4.5	68.2	18.2	9.1	366
	B	0.0	0.0	0.0	4.3	8.7	43.5	26.1	13.0	363
	C	0.0	0.0	0.0	0.0	9.9	81.8	9.1	0.0	351
36	A	0.0	0.0	0.0	0.0	0.0	77.8	11.1	11.1	375
	B	0.0	0.0	0.0	0.0	12.5	37.5	50.0	0.0	369
	C	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	400

(Rows indicate percentage of scores assigned at each stage on a given interview form at a given age.)

Table 12

Regression Coefficients and R^2 For Each Interview Form

Form	Constant	Coefficient (S.E.)	R^2
A	18.618	0.913 (.024)	.880
B	55.400	0.833 (.020)	.895
C	29.152	0.909 (.024)	.876

Table 13

Moral Judgment Stage Scores

	Subject and Testing Time	Standard Global Score Form A	Standard Global Score Form B	Standard Global Score Form C	Standard Global Score A&B&C Combined	Kramer Scores	Structural Issue Scores
	1-1	*1	2	2	1/2	-	-
	1-2	1/2	2	2/3	2	-	-
	1-3	2/3	*2	2/3	2/3	-	-
	1-4	2/3	3	2/3	3	-	-
	1-6	3/4	3/4		3/4	-	-
Construction case	2-1	1/2	1/2	1/2	1/2	2/4	1/3
	2-2	3	3	3/4	3	-	3/4
	2-3	3/4	3/4	3/4	3/4	3/4	4/5
	2-4	x3	x2/3	3/4	x3	x2/3	5
	2-5	4/5	4/5	4/5	4/5	-	-
	2-6	4/5	4/5	-	4/5	-	
	3-1	1/2	2	1/2	2	1/2	2
	3-2	2/3	2/3	2/3	2/3	-	2/3
	3-3	3/4	3	3/4	3/4	3/4	3
	3-4	3/4	3/4	x3	3/4	-	3/4
	3-5	3/4	3/4	-	3/4	-	4
	3-6	3/4	4	3/4	3/4	-	-

Table 13 (continued)

	4-1	2/3	2	2	2	2/3	2
	4-2	3	2/3	2/3	2/3	-	2
	4-3	3/4	2/3	3/4	3	3/4	3/4
	4-4	3/4	*2/3	x3	3	3/4	3/4
	4-5	3/4	3/4	3/4	3/4	-	3/4
	5-1	2	1/2	1/2	1/2	1/2	1
	5-2	2/3	2	1/2	2/3	-	2
	5-3	2/3	x1/2	x1	x1/2	-	2
	5-4	3	2	2/3	2/3	-	2/3
	5-5	x2/3	2	3	2/3	-	2/3
	6-1	-	1/2	1/2	1/2	-	1/2
	6-2	-	*2/3	2/3	2/3	-	2
	6-3	-	3	3	3	-	2/3
	6-4	-	3	x2/3	x2/3	-	2/3
	6-5	-	x*2/3	2/3	2/3	-	3/4
	8-1	2	1/2	1	1/2	-	1/2
	8-2	2	x1	2	2	-	1/2
	8-4	3/4	3	3	3	-	3
Construction case	9-1	1/2	1/2	1/2	1/2	1/2	1/2
	9-2	2/3	2/3	1/2	2/3	1/2	2
	9-3	2/3	3	3	3	3/4	3
	9-4	3	3	3/4	3	-	3/4
	9-5	3/4	3/4	3/4	3/4	-	3/4
	9-6	4	*3/4	-	4	-	-

Table 13 (continued)

11-1	1/2	2	1/2	1/2	-	-
11-2	2/3	*2	2/3	2/3	-	-
12-1	2	1/2	1/2	1/2	1/2	1/2
12-2	3	2	3	2/3	-	2
12-3	3/4	3/4	3	3/4	4	3
12-4	3/4	x3	3/4	3/4	3/4	3
13-1	1/2	2	1/2	2	-	-
13-2	2	2	2	2	-	-
13-3	3	3	3	3	-	-
14-1	2/3	2	2/3	2/3	-	1/2
14-2	2/3/4	2/3	3	2/3	-	2/3
14-4	4	4	4	4	-	4
14-5	-	4/5	-	4/5	-	4
15-1	2/3	2/3	2/3	2/3	-	-
15-2	2/3	3	2/3	2/3	-	-
16-1	2/3	2/3	2	2/3	-	1/2
16-2	3/4	3	2	2/3	-	2/3
16-3	3/4	3/4	3	3/4	-	3/4
16-4	4	4	3/4	4	-	4
16-6	4	4	-	4	-	-

Table 13 (continued)

on-	17-1	-	2	1/2	1/2	2/3	1/2
truc-	17-2	2/3	2/3	2/3	2/3	2/3	3/4
ion	17-3	3/4	3	3/4	3	3/4	3
ase	17-4	3/4	3	3/4	3/4	-	4
	17-5	4	4	3/4	4	-	4
	17-6	4/5	4/5	-	4/5	-	-
	18-1	3	3	2/3	3	-	2
	18-2	2/3	3	2/3	x2/3	-	3
	18-4	3/4	3/4	3/4	3/4	-	4
	18-5	3/4	*3/4	4	3/4	-	4/5
	18-6	4	4	-	4	-	-
	19-1	2	1/2	2/3	2/3	-	2
	19-2	2/3	3/4	3	3	-	2/3
	19-3	-	-	3/4	3/4	-	-
	19-4	3	3/4	3/4	3/4	-	2/3
	19-5	3/4	3/4	-	3/4	-	2/3
	19-6	3/4	4	-	4	-	-
	21-1	2	1/2	1/2	1/2	-	2
	21-2	2/3	2	2	2/3	-	-
	21-3	3	2/3	2/3	2/3	-	2/3
	21-4	x2/3	3	3	3	-	2/3

Table 13 (continued)

	22-1	2	1/2	2	2	1/2	2
	22-2	x1/2	2/3	x1/2	2/3	-	3
	22-3	3/4	3/4	3/4	3/4	2/4	3
	22-4	3/4	4	3/4	3/4	-	4
	22-5	3/4	-	3/4	3/4	-	-
	22-6	4	4/5	-	4	-	-
Construction case	23-1	2	1/2	1/2	1/2	-	1/2
	23-2	2/3	3/4	-	3/4	-	-
	23-3	3/4	3/4	3/4	3/4	-	4
	23-4	3/4	4/5	3/4	3/4	-	-
	23-5	4/5	5	4	4/5	-	4/5
	23-6	4/5	-	-	4/5	-	
	24-1	3	1/2	1/2	1/2	1/3	1/2
	24-2	3	2/3	2/3	2/3		2/3
	24-3	3/4	3	2/3	3	2/3	3
	24-4	3/4	3/4	3/4	3/4	2/4	3/4
	25-1	2/3	1/2	2/3	2/3	-	2/3
	25-2	3/4	3	2/3	3	-	2/3
	25-5	3/4	3/4	3/4	3/4	-	4/5
	25-6	3/4	4	-	4	-	

Table 13 (continued)

26-1	2	2	2/3	2	2/4	2
26-2	3	*2/3	3/4	3	3/4	3
26-4	3	3	x3	3	-	3
26-5	3	*3/4	3/4	3	-	3
26-6	3/4	3/4	-	3/4	-	-
27-1	2/3	1/3	2/3	2/3	3/4	3
27-2	3/4	3	3	3	3/4	3
27-4	3/4	3	3/4	3/4	-	3
27-5	3/4	3/4	3/4	3/4	-	3
29-1	2/3	2/3	2/3	2/3	3/4	-
29-2	3	3	3	3	2/4	-
29-4	3/4	3	3	3/4	-	-
29-5	4	3/4	3/4	3/4	-	-
31-1	1/2/3	2	2/3	2/3		-
31-2	3	2/3	3/4	3		-
31-5	3/4	3/4	3/4	3/4		-
32-1	1/2	2	2/3	2		-
32-2	2/3	2	2/3	2/3		-
32-5	3	3/4	3	3		-
32-6	3/4	3/4	-	3/4		-

Table 13 (continued)

36-1	2/3	1/2	1/2	1/2	1/2	-
36-2	x1/2/2	2	1/2	1/2	1/2	-
36-5	3/4	3	3/4	3/4		-
36-6	3/4	3/4	-	3/4		-
37-1	3	2/3	3	3	3/4	3
37-2	3/4	3/4	3	3/4	3/5	3/4
37-3	*3/4	3/4	3	3/4	3/4	3/4
37-4	4	*4	3/4	4	3/4	3/4
37-5	4/5	4/5	-	4/5		5
37-6	-	4/5	-	4/5		-
38-1	2/3	2	2/3	2/3	2/4	-
38-2	3/4	3/4	3	3	3/4	-
38-5	3/4	4	3/4	3/4		-
38-6	3/4	x3/4	-	3/4		-
39-1	2/3	2/3	2/3	2/3	3/4	2/3
39-2	3/4	3	3/4	3/4	4/5	3/4
39-3	3/4	3	3/4	3/4	2/3	4
39-4	4	3/4	3/4	3/4	4/5	4/5

Table 13 (continued)

	41-1	2/3	2	2	2	2/4	2
	41-2	2/3	2/3	3	3	3/4	3
	41-3	3	3	3/4	3	3/5	3/4
	41-4	3	3	3/4	3	3/4	3/4
	41-5	3	3	3/4	3	-	4
	41-6	3/4	4	-	4	-	-
Construction Case	42-1	2/3	2/3	3	2/3	2/4	2
	42-2	3	3	3/4	3/4	3/4	3
	42-3	-	-	4	4	3/5	3/4
	42-4	4/5	4	4	4	3/4	3/4
	42-5	4/5	4/5	4	4/5	-	4
	42-6	4/5	4/5	-	4/5	-	-
	43-1	*3	2/3	2/3	2/3	3	
	43-2	3/4	3/4	3	3/4	3/5	
	43-3	3/4	3/4	3	3/4	3/4	
	43-4	3/4	3/4	3/4	3/4	3/5	
	44-1	3/4	3/4	3/4	3/4	4/5	
	44-3	4/5	3/4	4	4	4/5	
	44-4	x4	4/5	x3/4	4	4/5	
	44-5	4/5	4/5	-	4/5	-	
	44-6	4/5	4/5	-	4/5	-	

Table 13 (continued)

45-1	2/3	2	2	2/3	2/4	2/3
45-2	3	2/3	2/3	2/3	3/4	2/3
45-4	4	3/4	3/4	3/4	3/4	3
45-5	4	4	4	4		4/5
45-6	4	4	-	4		-
47-1	2/3	3/4	2/3	2/3	2/3	-
47-2	2/3	3/4	3	3	2/4	2/3
47-3	3	-	x2/3	3	3/4	2/3
47-4	3	3/4	3	3/4	3/4	2/3
47-5	3/4	3/4	3	3/4		2/3
47-6	3/4	x2/4	-	4		-
48-1	2/3	-	2/3	-	-	-
48-2	3/4	2/3	3	3		23/
48-4	3/4	3	4	3/4		3/4
48-5	3/4	3/4	x3/4	3/4		-
48-6	3/4	4	-	4		-
49-1	3/4	3	3/4	3	-	-
49-4	3/4	3/4	3/4	3/4	-	-
49-5	3/4	x3	3/4	3/4	-	-
49-6	3/4	3/4	-	3/4	-	-

Table 13 (continued)

50-1	3	2/3	2	2/3		-
50-4	3/4	3	3	3		-
50-5	3/4	4	3/4	4		-
50-6	3/4	4	-	4		-
51-1	2	2	-	2		-
51-4	3/4	3/4	3	3/4		-
51-5	3/4	4	3/4	3/4		-
51-6	3/4	x3/4	-	3/4		-
53-1	3	*1/2/3	2	2/3	-	-
53-2	3/4	*3/4/5	2/3	3/4	-	-
53-4	x3	3/4	3/4	3/4	-	-
54-1	2/3	2/3	2	2/3		
54-5	4	3/4	3/4	3/4		
56-1	2	-	-	2		
56-4	3	-	-	3		
56-6	3/4	-	-	3/4		
59-1	2/3	1/2	2/3	2/3		
59-2	3	2/3	3	2/3		
59-4	3	3	3	3		
59-5	3/4	x2/3	3/4	3		

Table 13 (continued)

	62-1	2/3	3	3	3	
	62-2	3/4	4	3/4	3/4	
	62-4	x3	x*3	3/4	x3	
	62-6	3/4	4	-	4	
	64-1	3	2/3	3	3	
	64-4	3	3/4	3	3/4	
	64-5	3/4	3/4	3/4	3/4	
Construction Case	65-1	3	3	3	3	4/5
	65-2	4	-	x2/3	x2/3/4	2/4
	65-4	x3/4	3/4	3/4	3/4	-
	65-5	3/4	3/4	4	3/4	-
	65-6	3/4	4	-	4	-
Construction Case	67-1	3	2/3	2/3	2/3	3/4
	67-2	3	3	3/4	3	3/5
	67-3	-	-	-	-	4/5
	67-4	3/4	3/4	3/4	3/4	-
	67-5	4/5	3/4	3/4	4	-
	67-6	4/5	3/4	-	4/5	-

Table 13 (continued)

68-1	3	3	2	2/3
68-2	3/4	3	3	3/4
68-4	3/4	3/4	3	3/4
68-5	4	*3/4	3/4	3/4
68-6	4	x3	-	3/4
70-1	2/3	2/3	2/3	2/3
70-2	3/4	3/4	3/4	3/4
70-4	3/4	3/4	x3	3/4
70-5	4	3/4	3/4	3/4
70-6	x3/4	4	4	4
71-1	3	3	2/3	2/3
71-2	3/4	3	3	3
71-4	3/4	3/4	3	3/4
71-5	3/4	x3	3	3/4
81-1	3	1/2	2/3	2/3
81-2	3	2/3	3	3
81-5	3/4	3/4	3/4	3/4
91-3	3	3/4	3	3/4
91-4	x2/3	x3	3	x3
91-5	3/4	3	3/4	3
91-6	4	4	-	4

Table 13 (continued)

92-3	2/3	3	3/4	3
92-4	3/4	3/4	3/4	3/4
92-5	x3	*3/4	3/4	3/4
93-3	3/4	3	3	3
93-4	3/4	3	3	3
93-5	3/4	3/4	3	3/4
93-6	3/4	4		4
95-3	3	2/3	3	3
95-4	3/4	-	3	3/4
95-5	3/4	3	3/4	x3
95-6	-	3	-	3
96-3	3/4	3	2/3	3
96-4	x3	3	3	3
96-5	3/4	3/4	3/4	3/4
96-6	3/4	x3	-	x3

Table 14

Comparison of Downward Stage

Change: longitudinal* and test-retest

Pure and Mixed Stage Scores (9 point scale)

Longitudinal	Form A (Rater 1)	Form B (Rater 2)	Form C	Forms A, B, C Combined
$T_n \longrightarrow T_{n+1}$	5%	6%	4%	3%
Test-retest				
$T_1 \longrightarrow T_2$	19%	23%	No Data	No Data

Major/Minor and Pure Stage Scores (13 point scale)

Longitudinal	Form A	Form B	Form C	Forms A, B, C Combined
$T_n \longrightarrow T_{n+1}$	7%	6%	6%	5%
Test-retest				
$T_1 \dashrightarrow T_2$	19%	33%	No Data	No Data

*Only the longitudinal interviews which were scored blind are included in this analysis.

Table 15

Correlations Among 18 Moral Issues for Total Sample

	A1	A2	A3	A4	A5	A6
A2	.74 (204) P=0.00	1.00 (0) P=****	.67 (149) P=0.00	.76 (154) P=0.00	.66 (200) P=0.00	.63 (193) P=0.00
A3	.73 (165) P=0.00	.67 (149) P=0.00	1.00 (0) P=****	.68 (115) P=0.00	.61 (166) P=0.00	.61 (159) P=0.00
A4	.69 (156) P=0.00	.76 (154) P=0.00	.68 (115) P=0.00	1.00 (0) P=****	.65 (156) P=0.00	.59 (154) P=0.00
A5	.70 (219) P=0.00	.66 (200) P=0.00	.61 (166) P=0.00	.65 (156) P=0.00	1.00 (0) P=****	.66 (212) P=0.00
A6	.67 (290) P=0.00	.63 (193) P=0.00	.61 (159) P=0.00	.59 (154) P=0.00	.66 (212) P=0.00	1.00 (0) P=****

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	B1	B2	B3	B4	B5	B6
A2	.74 (174) P=0.00	.66 (167) P=0.00	.74 (161) P=0.00	.69 (170) P=0.00	.67 (165) P=0.00	.76 (139) P=0.00
A3	.74 (144) P=0.00	.70 (138) P=0.00	.73 (135) P=0.00	.66 (140) P=0.00	.52 (136) P=0.00	.70 (113) P=0.00
A4	.69 (136) P=0.00	.65 (128) P=0.00	.68 (121) P=0.00	.69 (135) P=0.00	.61 (131) P=0.00	.77 (103) P=0.00
A5	.68 (187) P=0.00	.69 (179) P=0.00	.71 (171) P=0.00	.64 (180) P=0.00	.62 (182) P=0.00	.67 (149) P=0.00
A6	.64 (178) P=0.00	.65 (175) P=0.00	.64 (170) P=0.00	.64 (171) P-0.00	.59 (175) P=0.00	.66 (144) P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	C1	C2	C3	C4	C5	C6
A2	.65 (115) P=0.00	.70 (77) P=0.00	.71 (152) P=0.00	.74 (132) P=0.00	.70 (161) P=0.00	.68 (159) P=0.00
A3	.57 (97) P=0.00	.55 (68) P=0.00	.60 (119) P=0.00	.67 (100) P=0.00	.60 (127) P=0.00	.62 (125) P=0.00
A4	.64 (87) P=0.00	.75 (58) P=0.00	.67 (113) P=0.00	.79 (104) P=0.00	.66 (121) P=0.00	.66 (117) P=0.00
A5	.62 (131) P=0.00	.60 (93) P=0.00	.55 (165) P=0.00	.61 (144) P=0.00	.62 (177) P=0.00	.62 (174) P=0.00
A6	.57 (129) P=0.00	.66 (86) P=0.00	.57 (160) P=0.00	.57 (141) P=0.00	.53 (172) P=0.00	.60 (167) P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	A1	A2	A3	A4	A5	A6
B1	.69	.74	.74	.69	.68	.64
	(188)	(174)	(144)	(136)	(187)	(178)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B2	.60	.66	.70	.65	.69	.65
	(179)	(167)	(138)	(128)	(179)	(175)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B3	.71	.74	.73	.68	.71	.64
	(176)	(161)	(135)	(121)	(171)	(170)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B4	.61	.69	.66	.69	.64	.63
	(178)	(.70)	(140)	(135)	(180)	(171)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B5	.50	.67	.52	.61	.62	.59
	(180)	(165)	(136)	(131)	(182)	(175)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B6	.66	.76	.70	.77	.67	.66
	(148)	(139)	(113)	(103)	(149)	(144)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	B1	B2	B3	B4	B5	B6
B1	1.00	.86	.88	.81	.71	.74
	(0)	(161)	(163)	(159)	(164)	(143)
	P=****	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B2	.86	1.00	.83	.83	.66	.78
	(161)	(0)	(152)	(161)	(157)	(131)
	P=0.00	P=****	P=0.00	P=0.00	P=0.00	P=0.00
B3	.88	.83	1.00	.79	.66	.74
	(163)	(152)	(0)	(148)	(156)	(135)
	P=0.00	P=0.00	P=****	P=0.00	P=0.00	P=0.00
B4	.81	.83	.79	1.00	.66	.72
	(159)	(161)	(148)	(0)	(158)	(128)
	P=0.00	P=0.00	P=0.00	P=****	P=0.00	P=0.00
B5	.71	.66	.66	.66	1.00	.83
	(164)	(157)	(156)	(158)	(0)	(133)
	P=0.00	P=0.00	P=0.00	P=****	P=0.00	P=0.00
B6	.74	.78	.74	.72	.83	1.00
	(143)	(131)	(135)	(128)	(133)	(0)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=****

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	C1	C2	C3	C4	C5	C6
B1	.66	.66	.68	.70	.63	.70
	(116)	(83)	(149)	(133)	(155)	(153)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B2	.62	.56	.67	.67	.62	.62
	(110)	(76)	(142)	(127)	(152)	(144)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B3	.68	.64	.66	.75	.64	.67
	(116)	(78)	(145)	(130)	(152)	(150)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B4	.72	.69	.64	.74	.61	.66
	(111)	(74)	(138)	(125)	(142)	(140)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B5	.56	.49	.60	.64	.52	.58
	(118)	(86)	(149)	(132)	(161)	(159)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
B6	.62	.55	.73	.77	.63	.64
	(102)	(74)	(119)	(108)	(127)	(130)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	A1	A2	A3	A4	A5	A6
C1	.65 (130) P=0.00	.65 (115) P=0.00	.57 (97) P=0.00	.64 (87) P=0.00	.62 (131) P=0.00	.57 (129) P=0.00
C2	.62 (91) P=0.00	.70 (77) P=0.00	.55 (68) P=0.00	.75 (58) P=0.00	.60 (93) P=0.00	.66 (86) P=0.00
C3	.62 (168) P=0.00	.71 (152) P=0.00	.60 (119) P=0.00	.67 (113) P=0.00	.55 (165) P=0.00	.57 (160) P=0.00
C4	.67 (147) P=0.00	.74 (132) P=0.00	.67 (100) P=0.00	.79 (104) P=0.00	.61 (144) P=0.00	.57 (141) P=0.00
C5	.59 (177) P=0.00	.70 (161) P=0.00	.60 (127) P=0.00	.66 (121) P=0.00	.62 (177) P=0.00	.53 (172) P=0.00
C6	.62 (175) P=0.00	.68 (159) P=0.00	.62 (125) P=0.00	.66 (117) P=0.00	.62 (174) P=0.00	.60 (167) P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	B1	B2	B3	B4	B5	B6
C1	.66	.62	.68	.72	.56	.62
	(116)	(110)	(116)	(111)	(118)	(102)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
C2	.66	.56	.64	.69	.49	.55
	(83)	(76)	(78)	(75)	(86)	(74)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
C3	.68	.67	.66	.64	.60	.73
	(149)	(142)	(145)	(138)	(149)	(119)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
C4	.70	.67	.75	.74	.64	.77
	(113)	(127)	(130)	(125)	(132)	(108)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
C5	.63	.62	.64	.61	.52	.63
	(155)	(152)	(152)	(142)	(161)	(127)
	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00	P=0.00
C6	.70	.62	.67	.66	.58	.64
	(153)	(144)	(150)	(140)	(159)	(130)
	P=0.00	P=0.00	P-.00	P=0.00	P=0.00	P=0.00

Table 15 (continued)

Correlations Among 18 Moral Issues for Total Sample

	C1	C2	C3	C4	C5	C6
C1	1.00 (0) P=****	.83 (75) P=0.00	.65 (120) P=0.00	.70 (101) P=0.00	.58 (129) P=0.00	.63 (127) P=0.00
C2	.83 (75) P=0.00	1.00 (0) P=****	.57 (86) P=0.00	.64 (71) P=0.00	.59 (89) P=0.00	.66 (91) P=0.00
C3	.65 (120) P=0.00	.57 (86) P=0.00	1.00 (0) P=****	.89 (146) P=0.00	.62 (167) P=0.00	.68 (164) P=0.00
C4	.70 (101) P=0.00	.64 (71) P=0.00	.89 (146) P=0.00	1.00 (0) P=****	.69)146) P=0.00	.75 (146) P=0.00
C5	.58 (129) P=0.00	.59 (89) P=0.00	.62 (167) P=0.00	.69 (146) P=****	1.00 (0) P=0.00	.82 (182) P=0.00
C6	.63 (127) P=0.00	.66 (91) P=0.00	.68 (164) P=0.00	.75 (146) P=0.00	.82 (183) P=0.00	1.00 (0) P=****

Table 16

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	A1	A2	A3	A4	A5	A6
A1	1.00 (0) P=****	.59 (37) P=0.00	.41 (26) P=0.02	.77 (29) P=0.00	.49 (42) P=0.00	.47 (37) P=0.00
A2	.58 (37) P=0.00	1.00 (0) P=****	.22 (23) P=0.16	.68 (27) P=0.00	.30 (37) P=0.04	.29 (33) P=0.05
A3	.41 (26) P=0.02	.22 (23) P=16	1.00 (0) P=0****	.11 (19) P=0.33	.11 (26) P=0.30	.25 (24) P=0.12
A4	.77 (29) P=0.00	.68 (27) P=0.00	.11 (19) P=0.33	1.00 (0) P=****	.52 (29) P-0.02	.21 (28) P=0.13
A5	.49 (42) P=0.00	.30 (37) P=0.03	.11 (26) P-0.30	.52 (29) P=0.00	1.00 (0) P=****	.41 (38) P=0.01
A6	.47 (37) P=0.00	.29 (33) P=0.05	.25 (24) P=0.12	.21 (28) P=0.34	.41 (38) P-0.01	1.00 (0) P=****

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	B1	**B2**	**B3**	**B4**	**B5**	**B6**
A1	.49 (37) P=0.00	.45 (34) P=0.00	.64 (30) P=0.00	.38 (30) P=0.02	.28 (36) P=0.5	.44 (28) P=0.01
A2	.57 (33) P=0.00	.47 (25) P=0.01	.50 (27) P=0.00	.70 (29) P=0.00	.46 (32) P=0.00	.71 (25) P=0.00
A3	.13 (24) P=0.27	.46 (25) P=0.01	.27 (19) P=0.13	.22 (20) P=0.18	.05 (24) P=0.41	.51 (19) P=0.01
A4	.64 (29) P=0.00	.35 (24) P=0.05	.57 (21) P=0.00	.61 (23) P=0.00	.37 (27) P=0.03	.49 (19) P=0.02
A5	.42 (37) P=0.01	.53 (34) P=0.00	.48 (20) P=0.00	.44 (30) P=0.01	.37 (37) P=0.01	.33 (29) P=0.04
A6	.26 (33) P=0.08	.46 (32) P=0.00	.59 (28) P=0.00	.36 (28) P=0.03	.10 (25) P=0.29	.22 (26) P=0.14

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	C1	C2	C3	C4	C5	C6
A1	.24 (28) P=0.11	.18 (19) P=0.23	.45 (33) P=0.00	.60 (28) P-0.00	.27 (37) P=0.05	.29 (36) P=0.05
A2	..41 (25) P=0.02	.46 (17) P=0.03	.59 (28) P=0.00	.57 (25) P=0.00	.54 (33) P=0.00	.50 (32) P=0.00
A3	.29 (21) P=0.10	-.07 (14) P=0.41	.19 (21) P=0.20	.33 (16) P=0.11	.07 (24) P=0.38	.19 (23) P=0.19
A4	.07 (19) P=0.39	.35 (13) P=0.12	.41 (22) P=0.03	.71 (22) P=0.00	.61 (26) P=0.00	.50 (24) P=0.01
A5	.21 (29) P=0.14	.31 (20) P=0.10	.17 (32) P=0.08	.37 (27) P=0.03	.47 (38) P=0.00	.49 (37) P=0.00
A6	.27 (27) P=0.09	.17 (19) P=0.25	.20 (29) P=0.15	.42 (25) P=0.02	.09 (34) P=0.31	.29 (33) P=0.05

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	A1	A2	A3	A4	A5	A6
B1	.49 (37) P=0.00	.57 (33) P=0.00	.13 (24) P=0.27	.64 (29) P=0.00	.42 (37) P=0.01	.26 (33) P=0.08
B2	.45 (34) P=0.00	.47 (39) P=0.01	.46 (25) P=0.01	.35 (24) P=0.05	.53 (34) P=0.00	.46 (32) P=0.00
B3	.64 (30) P=0.00	.50 (27) P=0.00	.27 (19) P=0.13	.57 (21) P=0.00	.48 (29) P=0.00	.59 (28) P=0.00
B4	.38 (30) P=0.02	.70 (29) P=0.00	.22 (20) P=0.08	.61 (23) P=0.00	.44 (30) P=0.01	.36 (28) P=0.03
B5	.28 (36) P=0.05	.46 (32) P=0.00	.05 (24) P=0.41	.37 (27) P=0.03	.37 (37) P=0.01	.10 (35) P=0.29
B6	.44 (28) P=0.01	.71 (25) P=0.00	.51 (19) P=0.01	.49 (19) P=0.02	.33 (29) P=0.04	.22 (26) P=0.14

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	B1	B2	B3	B4	B5	B6
B1	1.00 (0) P=****	.67 (32) P=0.00	.74 (28) P=0.00	.78 (27) P=0.00	.45 (34) P=0.00	.41 (26) P=0.02
B2	.67 (32) P=0.00	1.00 (0) P=****	.56 (26) P=0.00	.63 (26) P=0.00	.38 (32) P=0.02	.24 (25) P=0.13
B3	.74 (28) P=0.00	.56 (26) P=0.00	1.00 (0) P=****	.77 (21) P=0.00	.14 (28) P=0.25	.39 (19) P=0.05
B4	.78 (27) P=0.00	.63 (26) P=0.00	.77 (21) P=0.00	1.00 (0) P=****	.69 (27) P=0.00	.50 (20) P=0.01
B5	.45 (34) P=0.00	.38 (31) P=0.02	.14 (28) P=0.25	.69 (27) P=0.00	1.00 (0) P=****	.55 (25) P=0.00
B6	.41 (26) P=0.02	.24 (25) P=0.13	.39 (19) P=0.05	.50 (20) P=0.01	.55 (25) P=0.00	1.00 (0) P=****

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	C1	C2	C3	C4	C5	C6
B1	.29 (25) P=0.08	.57 (19) P=0.01	.31 (30) P=0.05	.54 (27) P=0.00	.43 (34) P=0.01	.52 (33) P=0.00
B2	.23 (26) P=0.13	.35 (18) P=0.08	.34 (28) P=0.04	.36 (24) P=0.04	.26 (33) P=0.07	.55 (31) P=0.00
B3	.55 (21) P=0.01	.57 (14) P=0.02	.38 (25) P=0.03	.50 (23) P=0.01	.21 (27) P=0.15	.46 (27) P=0.01
B4	.49 (22) P=0.01	.68 (16) P=0.00	.43 (26) P=0.01	.68 (21) P=0.00	.48 (27_ P=0.01	.62 (26) P=0.00
B5	.18 (26) P=0.09	.24 (20) P=0.06	.18 (28) P=0.18	.23 (25) P=0.14	.21 (35) P=0.11	.38 (35) P=0.02
B6	.03 (23) P=0.45	-.05 (16) P=0.43	.40 (22) P=0.03	.55 (19) P=0.01	.38 (27) P=0.03	.37 (27) P=0.03

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	A1	A2	A3	A4	A5	A6
C1	.24 (28) P=0.11	.41 (25) P=0.02	.29 (21) P=0.10	.07 (19) P=0.39	.21 (29) P=0.14	.27 (27) P=0.09
C2	.18 (19) P=0.23	.46 (17) P=0.03	-.06 (14) P=0.41	.35 (13) P=0.12	.31 (20) P=0.10	.17 (19) P=0.25
C3	.45 (33) P=0.00	.59 (28) P=0.00	.19 (21) P=0.20	.41 (22) P=0.03	.17 (32) P=0.18	.20 (29) P=0.15
C4	.60 (28) P=0.00	.57 (25) P=0.00	.33 (16) P=0.11	.71 (22) P=0.00	.37 (27) P=0.03	.42 (25) P=0.12
C5	.27 (37) P=0.05	.54 (33) P=0.00	.07 (24) P=0.38	.61 (26) P=0.00	.47 (38) P=0.00	.09 (34) P=0.31
C6	.29 (36) P=0.05	.50 (32) P=0.00	.19 (23) P=0.19	.50 (24) P=0.01	.49 (37) P=0.00	.29 (33) P=0.05

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	B1	B2	B3	B4	B5	B6
C1	.29 (25) P=0.08	.23 (26) P=0.13	.55 (21) P=0.01	.49 (22) P=0.01	.18 (26) P=0.19	.10 (23) P=0.45
C2	.57 (19) P=0.01	.35 (18) P=0.08	.57 (14) P=0.02	.68 (16) P=0.00	.24 (20) P=0.16	.05 (16) P=0.43
C3	.31 (30) P=0.05	.34 (28) P=0.04	.38 (25) P=0.03	.43 (26) P=0.01	.18 (28) P=0.09	.40 (22) P=0.03
C4	.54 (27) P=0.00	.36 (24) P=0.04	.50 (23) P=0.01	.68 (21) P=0.00	.23 (25) P=0.14	.55 (19) P=0.01
C5	.43 (34) P=0.01	.26 (33) P=0.07	.21 (27) P=0.15	.48 (27) P=0.01	.21 (35) P=0.11	.38 (27) P=0.03
C6	.52 (33) P=0.00	.56 (31) P=0.00	.46 (27) P=0.01	.62 (26) P=0.00	.38 (33) P=0.02	.37 (27) P=0.03

Table 16 (continued)

Correlations Among 18 Moral Issues for 16-17 year old Age Group

	C1	C2	C3	C4	C5	C6
C1	1.00 (0) P=****	.75 (17) P=0.00	.43 (25) P=0.02	.01 (19) P=0.48	.28 (28) P=0.08	.22 (29) P=0.12
C2	.75 (17) P=0.00	1.00 (0) P=****	.47 (17) P=0.03	.34 (12) P=0.14	.67 (18) P=0.00	.60 (19) P=0.00
C3	.43 (25) P=0.02	.47 (17) P=0.03	1.00 (0) P=****	.87 (26) P=0.00	.49 (31) P=0.00	.58 (30) P=0.00
C4	.01 (19) P=0.48	.34 (12) P=0.14	.87 (26) P=0.00	1.00 (0) P=****	.59 (27) P=0.00	.69 (27) P=0.00
C5	.28 (28) P=0.09	.67 (18) P=0.00	.49 (31) P=0.00	.59 (27) P=0.00	1.00 (0) P=****	.77 (37) P=0.00
C6	.22 (29) P=0.02	.60 (19) P=0.00	.58 (30) P=0.00	.69 (27) P=0=00	.77 (37) P=0.00	1.00 (0) P=****

Table 17

Eigenvalues and Associated Percent of Variance for 18 Issues and for Each Form in the Total Sample and 3 Age Groups

	(A, B, C)	1st Eigenvalue	% Variance	2nd Eignevalue	% Variance
All S's	18 issues	12.35	68.6	0.84	4.7
N=190	6 issues A	4.35	72.5	0.47	7.8
	6 issues B	4.83	80.6	0.49	8.3
	6 issues C	4.43	73.8	0.63	10.4
16-18 yrs	18 issues	8.27	46.0	1.87	10.4
N=40	6 issues A	3.06	50.9	1.30	17.1
	6 issues B	3.69	61.4	1.05	17.5
	6 issues C	3.63	60.6	1.29	21.5
20-22 yrs	18 issues	7.91	43.9	2.05	11.4
N=29	6 issues A	2.57	42.8	1.16	19.4
	6 issues B	3.90	65.0	0.99	16.5
	6 issues C	3.26	54.3	1.12	18.6
28-30 yrs	18 issues	7.01	58.4	1.49	12.4
N=30	6 issues A	3.50	58.4	0.86	14.4
	6 issues B	4.30	71.7	0.74	12.3

Table 18

First Factor Loadings

N = 190

	18 x 18		6 x 6
Life	.80		.87
Law	.86		.86
Morality and Conscience	.79	Form A	.80
Punishment	.84	(Alone)	.82
Contract	.78		.79
Authority	.75		.76
Life	.89		.92
Law	.84		.91
Morality and Conscience	.88	Form B	.90
Punishment	.85	(Alone)	.87
Contract	.75		.79
Authority	.86		.86
Life	.79		.81
Law	.77		.78
Morality and Conscience	.80	Form C	.83
Punishment	.87	(Alone)	.90
Contract	.78		.79
Authority	.81		.86

Table 19

Unrotated Loadings for Five Factors in the 16-18 Age Group

	1	2	3	4	5
Life	.68	.34	.24	.19	-.14
Law	.79	.09	-.10	-.04	.26
Morality and Conscience	.33	.29	.29	.24	.29
Punishment	.76	.24	-.17	-.09	-.27
Contract	.58	.05	.11	-.16	-.28
Authority	.46	.08	.41	.20	-.16
Life	.77	-.08	.12	-.26	-.15
Law	.65	.01	.30	-.11	-.09
Morality and Conscience	.77	-.12	.46	.12	-.16
Punishment	.88	-.15	.08	-.28	.12
Contract	.51	.10	-.03	-.59	.28
Authority	.61	.54	-.07	-.12	.38
Life	.46	-.60	.30	.20	.38
Law	.64	-.75	-.12	.01	-.03
Morality and Conscience	.66	-.03	-.32	.48	.22
Punishment	.80	.29	-.31	.33	-.12
Contract	.65	-.18	-.53	.01	-.10
Authority	.74	-.12	-.29	-.01	-.11
Eignevalues	8.27	1.87	1.65	1.34	1.19
% of Variance	46.00	10.40	9.20	7.20	6.60

Table 20

Percentage of Subjects of Each Age Group at Each Stage

Age	10	13-14	16-18	20-22	24-26	28-30	32-33	36
Global Stage								
1/2	47.6	8.1	2.2	0	0	0	0	0
2	33.3	16.2	11.1	0	0	0	0	0
2/3	14.3	56.8	17.8	9.4	8.0	0	0	0
3	4.8	16.2	44.4	31.3	12.0	16.2	8.7	0
3/4	0	2.7	24.4	40.6	48.0	51.4	47.8	44.4
4	0	0	0	18.8	16.0	18.9	30.4	44.4
4/5	0	0	0	0	16.0	13.5	13.0	11.1
Mean MMS	189	246	290	327	357	361	369	375
S.D. MMS	31.4	37.7	43.7	36.1	49.1	42.3	41.5	25.6
N	21	37	46	33	25	38	23	9

Table 21

Correlations Among Moral Maturity Scores at Different Ages

AGE	10	13-14	16-18	20-22	24-26	28-30	32-33	36
10		0.395*	0.456*	.473*	0.198	-0.249		
(N)		(21)	(15)	(18)	(13)	(10)		
13-14	0.395*		0.699*	0.464*	0.696*	0.668*	0.571*	
(N)	(21)		(30)	(24)	(23)	(23)	(10)	
16-18	0.456	0.699*		0.682*	0.732*	0.685*	0.448*	0.164
(N)	(15)	(30)		(31)	(21)	(36)	(23)	(10)
20-22	0.473*	0.464*	0.638*		0.710*	0.481	0.764*	-0.119
(N)	(18)	(24)	(31)		(19)	(23)	(12)	(6)
24-26	0.198	0.696*	0.732*	0.710*		0.905*	0.862*	
(N)	(13)	(23)	(21)	(19)		(16)	(7)	
28-30	-0.249	0.668*	0.685*	0.481*	0.905*		0.809*	0.083
(N)	(10)	(23)	(36)	(23)	(16)		(21)	(10)
32-33		0.571*	0.448*	0.764*	0.862*	0.809*		0.880*
(N)		(10)	(23)	(12)	(7)	(21)		(8)
36			0.164	-0.119		0.083	0.880*	
(N)			(10)	(6)		(10)	(8)	

*p .05

Table 22

Correlations between Moral Judgment and Social Class, Sociometric Status, IQ, and Education

Age Group	10	13-14	16-18	20-22	24-26	28-30	32-33	36
N	21	36	46	34	24	38	23	10
$r_{MMS, SES}$	.60*	.42*	.38*	.54*	.36*	.22*	.41*	.57*
$r_{MMS, SMS}$	-.36	-.12	-.17	-.04	-.17	-.22	-.18	-.35
$r_{IQ, MMS}$	.19	.25	.17	.27	.37*	.51*	.60*	.37
$r_{Education, MMS}$					.69*	.54*	.59*	.77*

*p .05

Table 23

The Relation Between Education and Moral Judgment

Divided by Social Class

Working Class

Stage	Finished High School Only	Some College	Four Years College and/or BA	Graduate School MA and/or PhD
4/5				33% (N = 1)
4		17% (N = 1)	33% (N = 1)	33% (N = 1)
3/4	75% (N = 3)	83% (N = 5)	67% (N = 2)	33% (N = 1)
3	25% (N = 1)			

(Data taken from age groups 28-36)

Table 23 (continued)

The Relation Between Education and Moral Judgment

Divided by Social Class

Middle Class

Stage	Finished High School Only	Some College	Four Years College and/or BA	Graduate School MA and/or PhD
4/5			14% (N = 1)	50% (N = 5)
4		33% (N = 1)	57% (N = 4)	50% (N = 5)
3/4	50% (N = 2)	67% (N = 2)	29% (N = 2)	
3	50% (N = 2)			

(Data taken from age groups 28-36)

Figure 1

Mean Percentage of Reasoning at Each Stage for Each Age Group

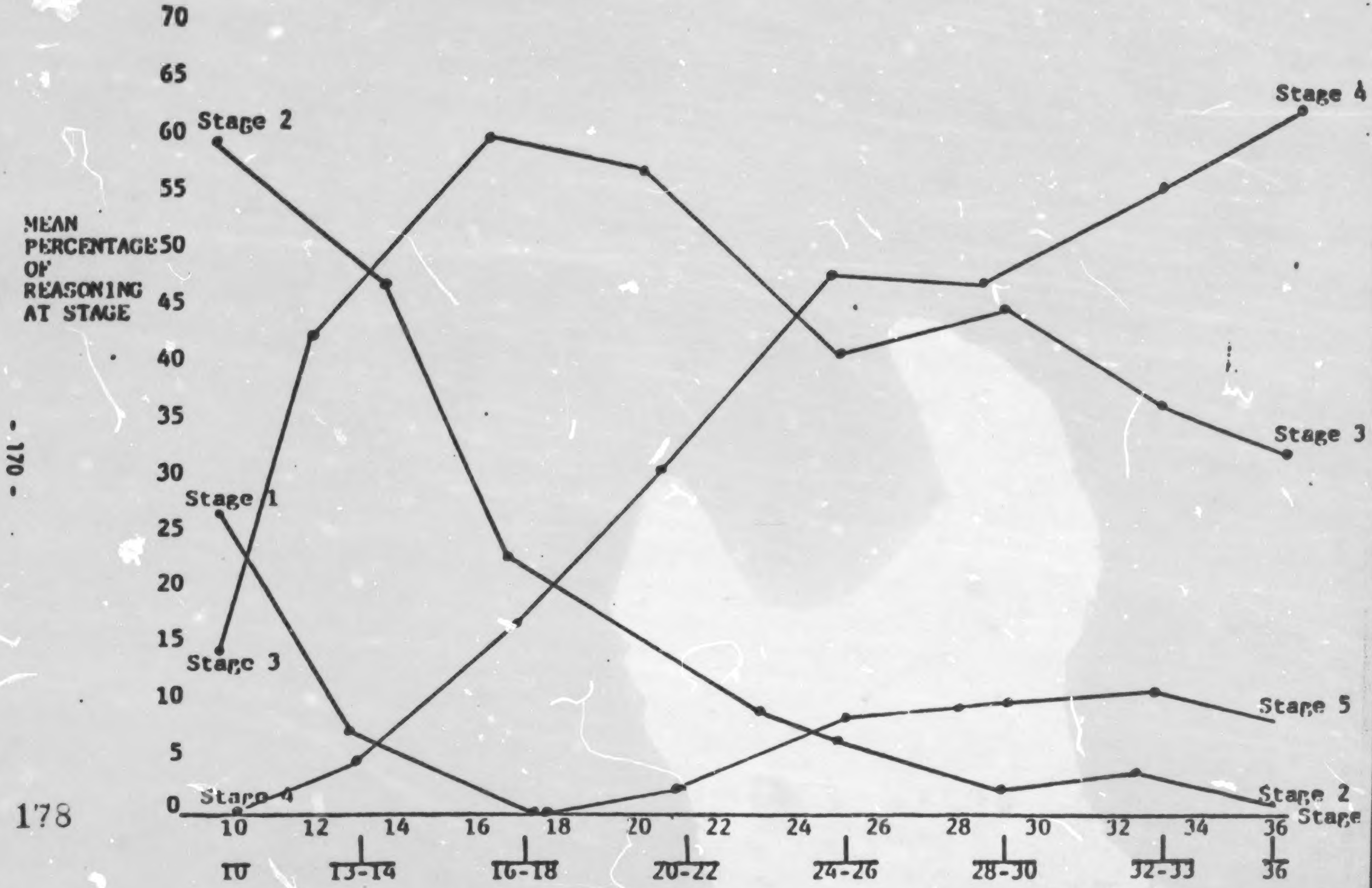

Figure 2

Percentage of Reasoning at Each Stage for Each Age

Subject 3 (10 years old at Time 1)

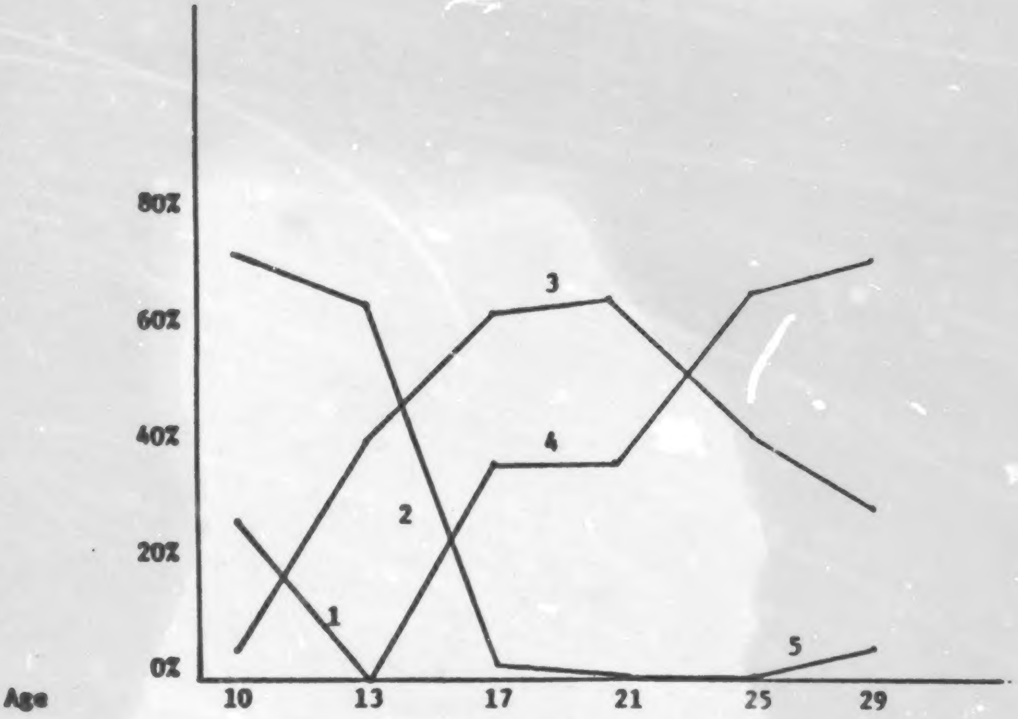

Figure 3

Percentage of Reasoning at Each Stage for Each Age

Subject 9 (10 years old at Time 1)

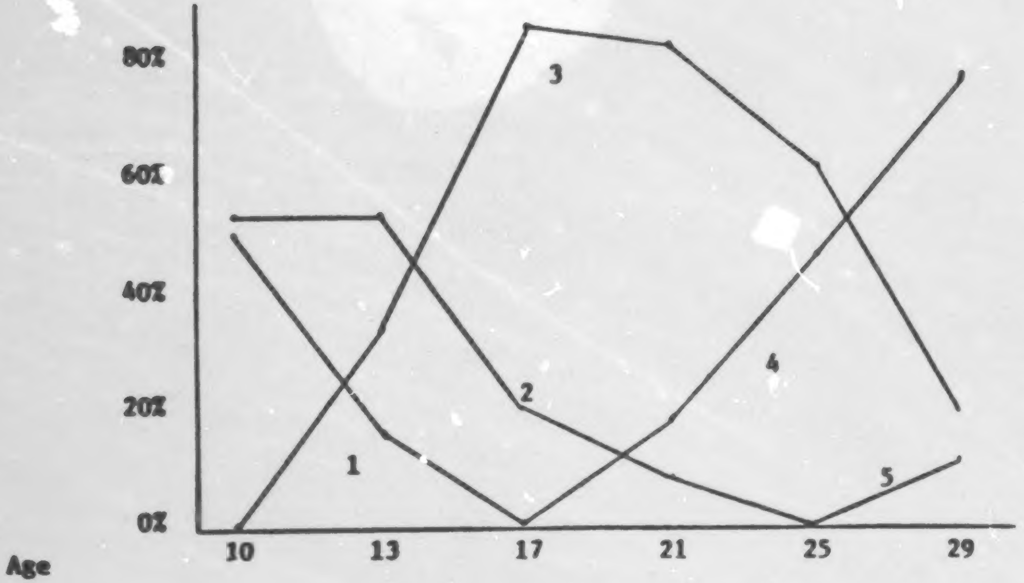

Figure 4

Percentage of Reasoning at Each Stage for Each Age

Subject 37 (13 years old at Time 1)

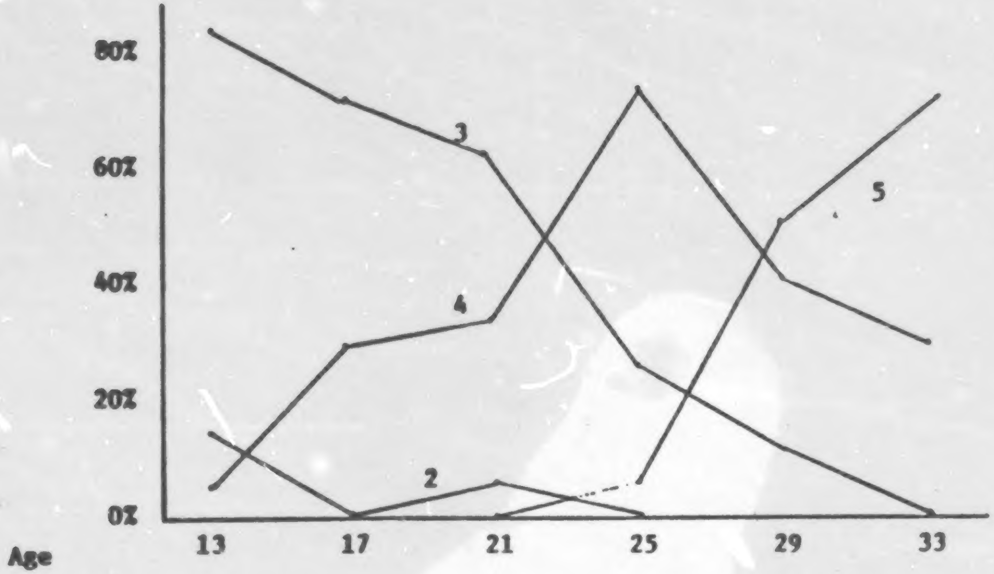

Figure 5

Percentage of Reasoning at Each Stage for Each Age

Subject 42 (13 years old at Time 1)

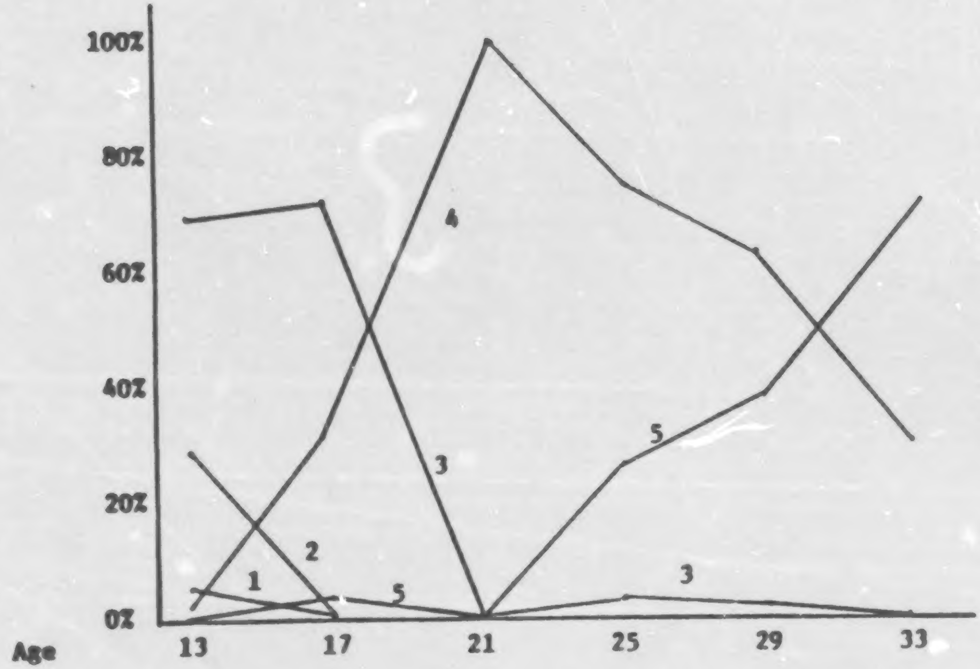

Figure 6

Mean Moral Maturity Scores at Each Age for Each Cohort

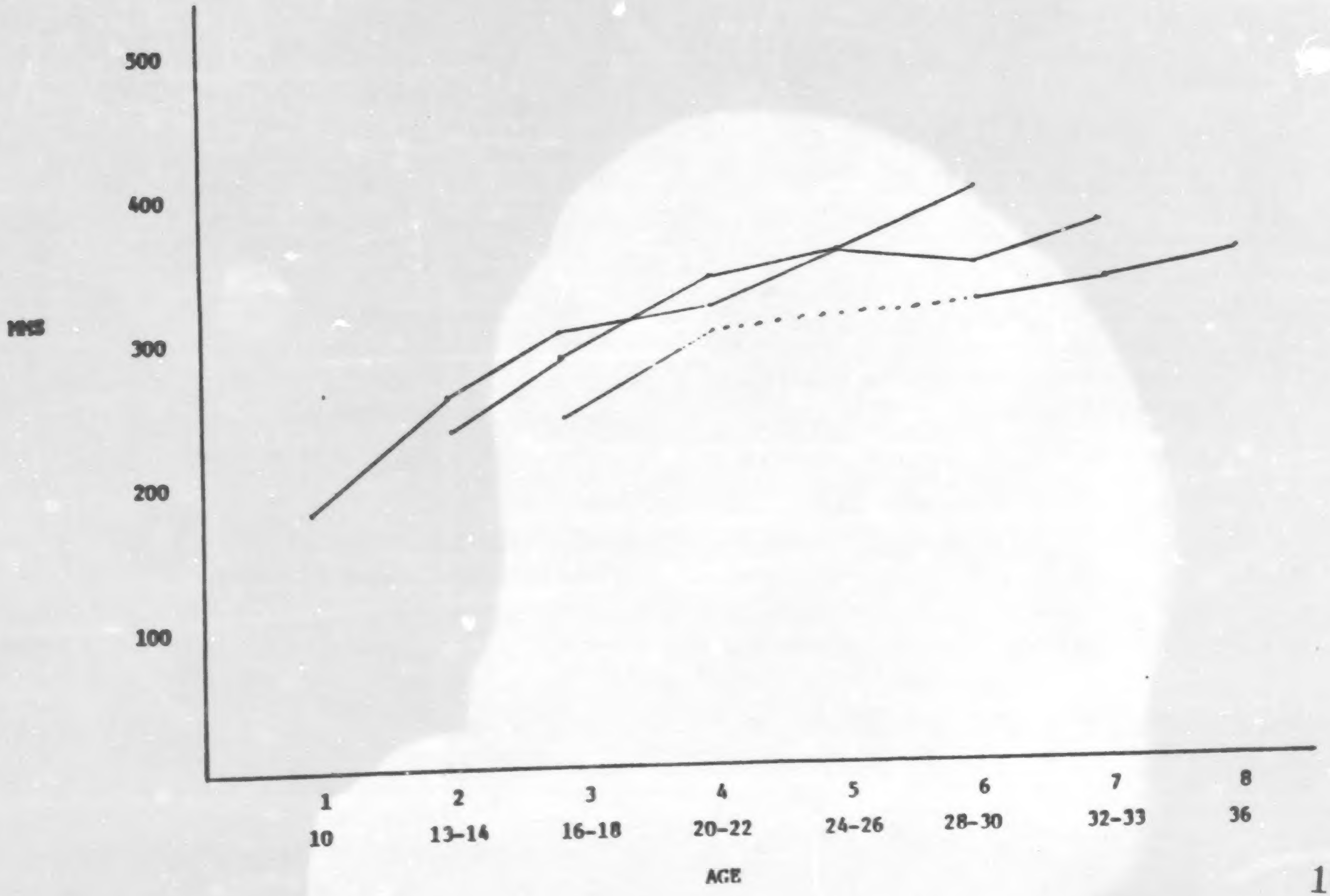

Figure 7a

Mean Percentage of Reasoning at Each Stage for Each Age Group - High SES Subject

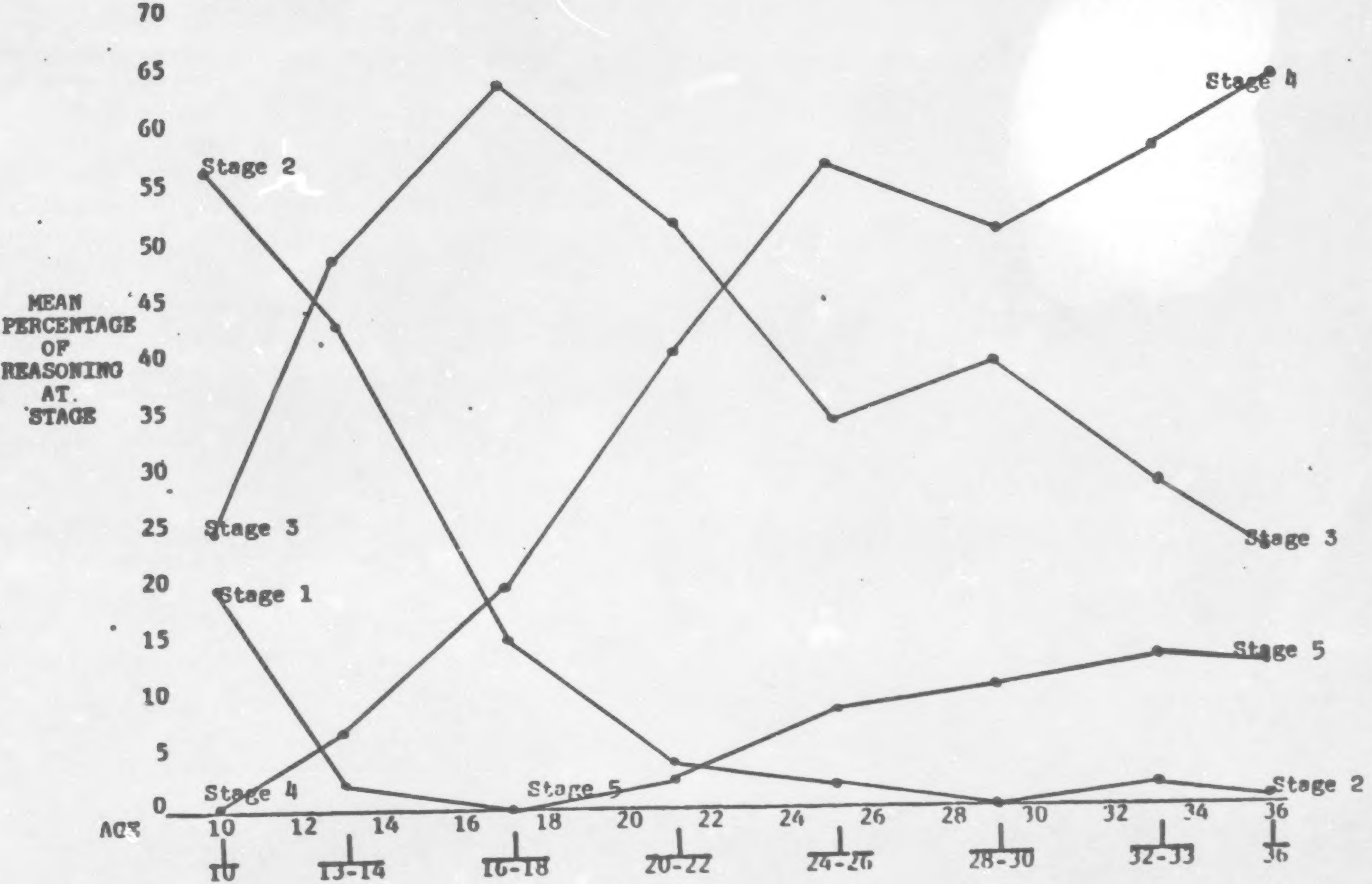

Figure 7b

Mean Percentage of Reasoning at Each Stage for Each Age Group - Low SES Subject

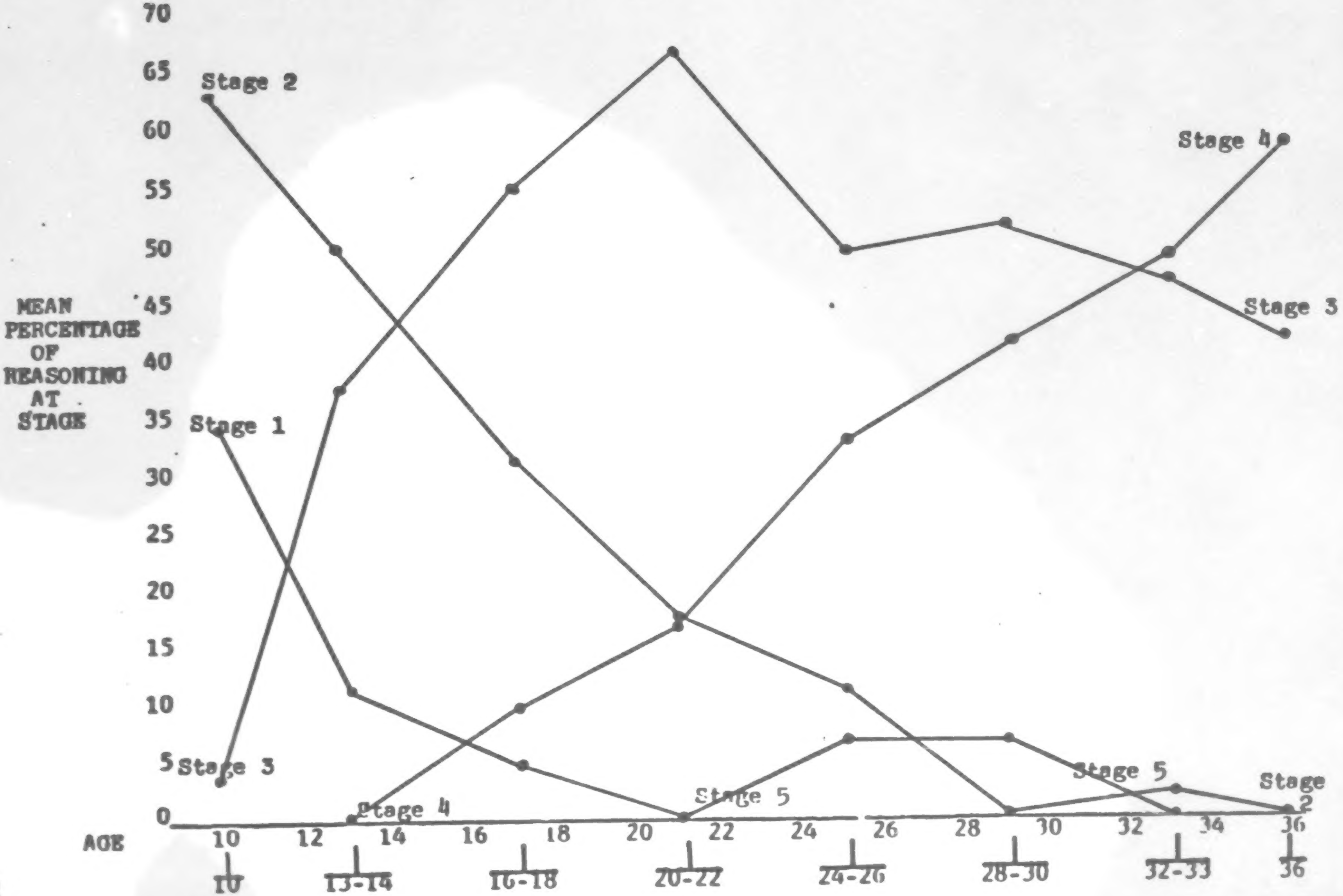

Figure 8a

Mean Percentage of Reasoning at Each Stage for Each Age Group - Sociometric Integrates

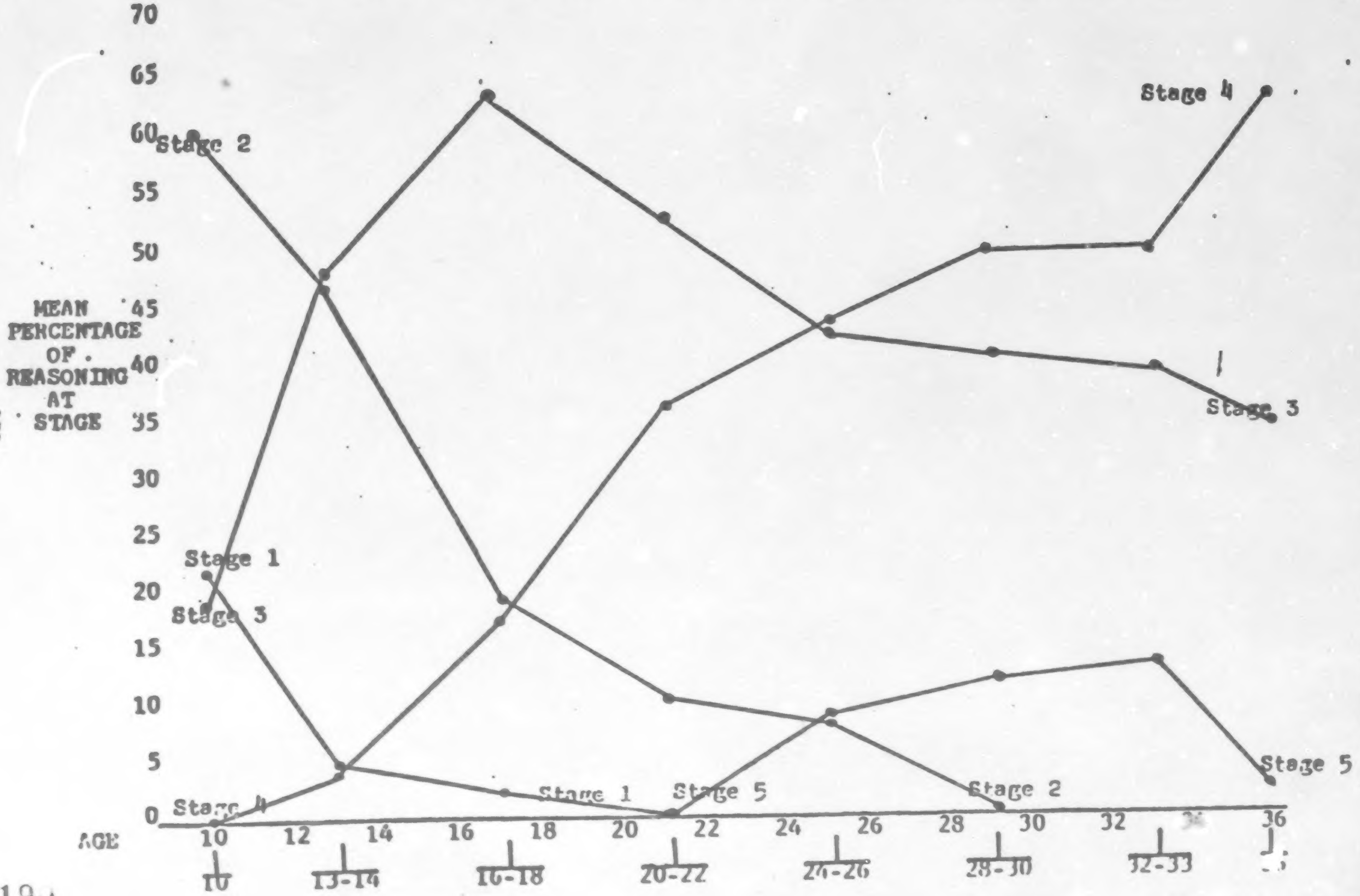

Figure 8b

Mean Percentage of Reasoning at Each Stage for Each Age Group - Sociometric Isolates

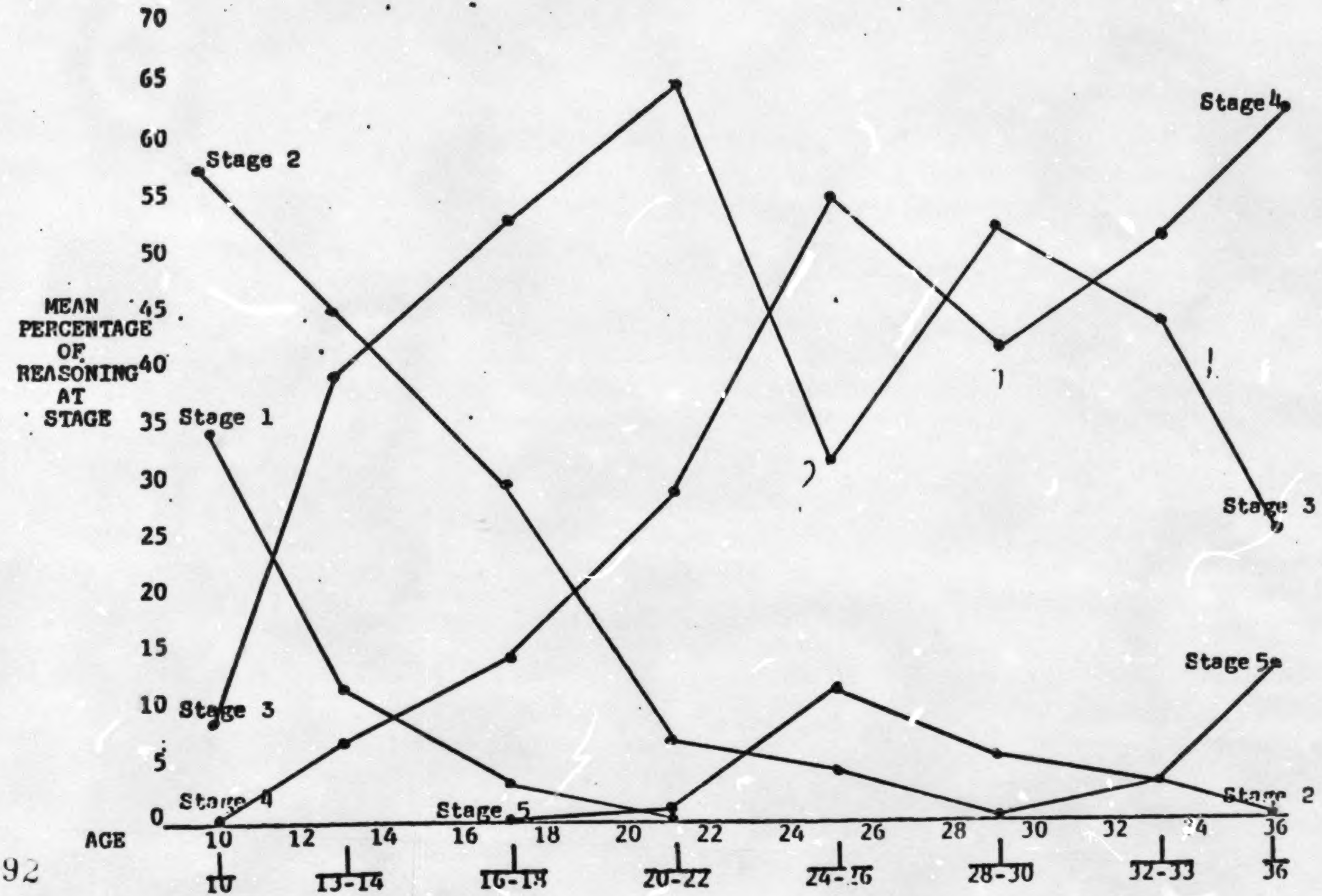

Figure 9

Distributions of Stage Usage in Relation to Modal Stage: Sentence Rating and Standard Scores

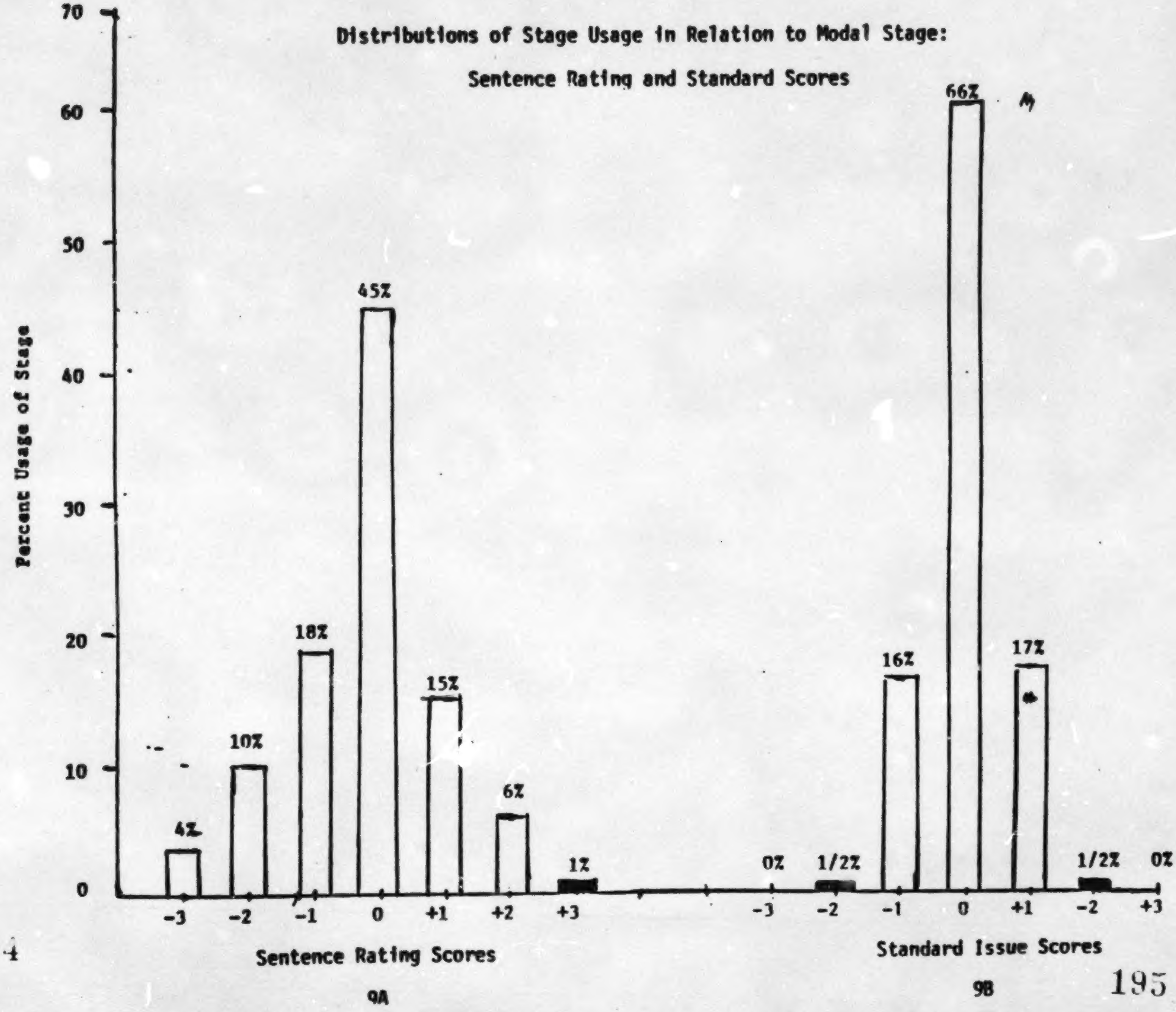

Figure 10a

Distribution of Stage Usage - Sentence Rating for High SES Subjects

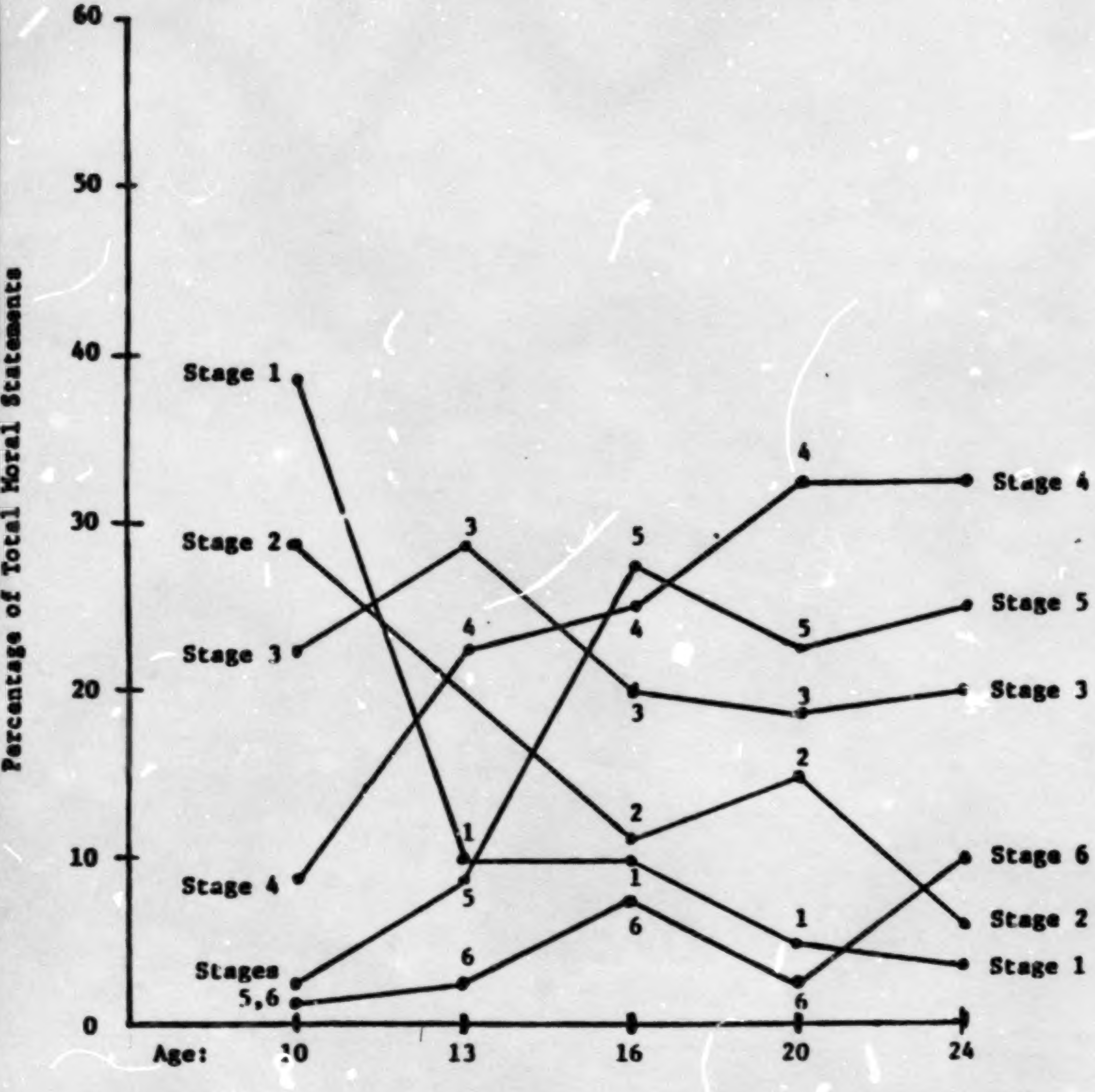

Figure 10b

Distribution of Stage Usage - Standard Scoring for High SES Subjects